

Garrett, Daniel; Pavan, Alessandro

**Working Paper**

## Managerial turnover in a changing world

Discussion Paper, No. 1490

**Provided in Cooperation with:**

Kellogg School of Management - Center for Mathematical Studies in Economics and Management Science, Northwestern University

*Suggested Citation:* Garrett, Daniel; Pavan, Alessandro (2010) : Managerial turnover in a changing world, Discussion Paper, No. 1490, Northwestern University, Kellogg School of Management, Center for Mathematical Studies in Economics and Management Science, Evanston, IL

This Version is available at:

<https://hdl.handle.net/10419/59621>

**Standard-Nutzungsbedingungen:**

Die Dokumente auf EconStor dürfen zu eigenen wissenschaftlichen Zwecken und zum Privatgebrauch gespeichert und kopiert werden.

Sie dürfen die Dokumente nicht für öffentliche oder kommerzielle Zwecke vervielfältigen, öffentlich ausstellen, öffentlich zugänglich machen, vertreiben oder anderweitig nutzen.

Sofern die Verfasser die Dokumente unter Open-Content-Lizenzen (insbesondere CC-Lizenzen) zur Verfügung gestellt haben sollten, gelten abweichend von diesen Nutzungsbedingungen die in der dort genannten Lizenz gewährten Nutzungsrechte.

**Terms of use:**

*Documents in EconStor may be saved and copied for your personal and scholarly purposes.*

*You are not to copy documents for public or commercial purposes, to exhibit the documents publicly, to make them publicly available on the internet, or to distribute or otherwise use the documents in public.*

*If the documents have been made available under an Open Content Licence (especially Creative Commons Licences), you may exercise further usage rights as specified in the indicated licence.*

# Managerial Turnover in a Changing World\*

**Daniel Garrett**

Northwestern University  
dfgarrett@gmail.com

**Alessandro Pavan**

Northwestern University  
alepavan@northwestern.edu

First version: December 2009. This version: May 2010

## Abstract

We characterize a firm's profit-maximizing turnover policy in an environment where managerial productivity changes stochastically over time and is the managers' private information. Our key positive result shows that the productivity level that the firm requires for retention declines with the managers' tenure in the firm. Our key normative result shows that, compared to what is efficient, the profit-maximizing policy 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.

*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, Marco Ottaviani, Bill Rogerson, and seminar participants at Berkeley Haas, Stanford, Northwestern, the European University Institute, and the 2010 Econometric Society Winter Meetings.

# 1 Introduction

The job security and pay of a firm's top managers 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.<sup>1</sup> As a result, the contracts that successful firms offer to their top managers are designed not only to incentivize their effort but also to guarantee the desired level of turnover. However, managers typically have better 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.<sup>2</sup> 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 management and for leaving the firm when these prospects deteriorate (equivalently, when the quality of their match with the firm is not satisfactory anymore).

While the literature on managerial contracting has focused primarily on compensation, turnover is also believed to play a key role in organizations, as is clear from the vast attention it typically receives in the press. However, the interaction between compensation and turnover in environments in which managerial productivity (equivalently, the match quality) is expected to change over time and to be privately observed by the managers remains largely unexplored.

The purpose of this paper is to provide a tractable, yet rich, model of managerial compensation which, in addition to the familiar role of incentivizing effort, accounts explicitly for the following possibilities: (i) a manager's ability to generate profits changes (stochastically) over time; (ii) shocks to a manager's productivity are anticipated but to a large extent privately observed by the manager; (iii) the board can respond to poor future prospects by replacing the incumbent manager with a new one; and (iv) the firm's performance under a new manager is affected by the same incentive and

---

<sup>1</sup>See, for example, Fine (1998), who argues that technology is increasing the speed at which business environments evolve across a plethora of industries.

<sup>2</sup>That the productivity of a worker in a particular job might depend on the quality of his match in that job, in addition to his innate ability, has been viewed as an important possibility at least since the seminal work of Jovanovic (1979).

informational problems as with the incumbent. Accounting for these possibilities not only is realistic, it permits us to shed new light on the dynamics of effort, performance, and retention.

**Model Preview.** A firm’s board of directors (the principal) hires a succession of managers (the agents) to operate the firm. 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 an incumbent manager, the board goes back to the labor market and is randomly matched with a new manager.<sup>3</sup> 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 meant to capture how the interaction of the environment with the tasks 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 calendar time plays no role. As a result, the board offers the same contract to each manager it hires. A contract is conveniently described in terms of: (i) the effort policy it induces—this policy specifies the effort the manager is recommended to exert as a function of the evolution of his productivity; (ii) the compensation the manager receives over time as a function of observed cash flows; and (iii) a turnover policy which specifies under which conditions separation will occur.

The positive and normative properties of the dynamics of effort, performance, and turnover are identified by characterizing the contract that maximizes the firm’s expected profits (net of managerial compensation) and comparing it to the contract that a benevolent planner would offer to each manager to maximize 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

---

<sup>3</sup>While the focus of the paper is the turnover induced by the contract that the firm offers to its managers, we do not model explicitly the firm’s search effort in the managerial labor market, after separation occurs. Instead, in the spirit of the search and matching literature, we assume that the results of such effort are captured by a stochastic process whose properties we describe in detail below.

by comparing, after each history, the value of continuing the relationship with an incumbent manager (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 the incumbent is hired. Given the stationarity of the environment, the payoff from hiring a new manager coincides with the payoff expected when hiring the incumbent manager. Both the profit-maximizing and the efficient policies are thus obtained as the solution to a recursive dynamic-programming problem and can be conveniently described as a fixed point to a particular functional mapping which summarizes all relevant trade-offs.

**Key positive results.** Our key positive prediction is that firms' retention policies become *gradually more lenient over time*: the productivity level required for a manager to be retained declines with the number of periods the manager has been working for the firm. 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;<sup>4</sup> 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 of generating the same distribution of present and future cash flows as less productive managers by working less, thus economizing on the disutility of effort.<sup>5</sup> Contrary to Laffont and Tirole's static setting, in our dynamic environment firms have two instruments to limit such rents: first, they can induce less productive types to work less (e.g., by offering them contracts with low-powered incentives such as payments that are relatively insensitive to realized cash flows); in addition, they can commit to a turnover policy that is more severe to a manager whose initial productivity is low in terms of the future productivity levels required to be retained. Both instruments play the role of discouraging managers who are initially most productive from mimicking less productive ones and are thus most effective when targeted at managers whose initial productivity is low.

---

<sup>4</sup>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.

<sup>5</sup>Equivalently, by the possibility of generating higher cash flows for the same amount of effort.

The key observation is that, when the effect of the managers' initial productivity on their subsequent productivity is expected to decline over time, then the effectiveness of such instruments is higher when these instruments are used at the early stages 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 managerial productivity.

The optimal retention policy is then driven by two considerations. First, the desire to respond efficiently to variations in managerial productivity, of course taking into account future effort and retention decisions. This concern calls for retaining managers whose productivity is expected to remain sufficiently high irrespective of whether their initial productivity was low. Second, the value of committing to a policy that reduces the rents that the firm must leave to the managers who are most productive at the initial contracting stage. As explained above, this second concern calls for a more severe retention policy for a manager 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 also provides a possible explanation for why managers with a longer tenure eventually become entrenched, in the sense that they are retained under the same conditions that would have called for separation at a shorter tenure.<sup>6</sup> Note that our explanation is fundamentally different from the alternative view that managers with longer tenure are able to exert more influence over the board and thus become less willing to adapt (e.g., Miller, 1991, and Rose and Shepard, 1997).<sup>7</sup>

A possible difficulty in testing for our key positive finding is that neither managerial productivity, nor the quality of the match between the firm and the managers, are directly observable. Furthermore, it is difficult to infer them from correlation with observed cash flows. This is true even when cash flows are a deterministic function of effort and managerial productivity (that is, in the absence of noise). The reason is that, under the optimal contract, effort typically increases over time, implying that the cash flows that the manager must generate to be retained need not be decreasing in the

---

<sup>6</sup>Note, however, that the relationship between tenure and *average* productivity over the relevant period may be ambiguous due to composition effects (see the discussion at the end of Section 4).

<sup>7</sup>In a similar vein, Shleifer and Vishny (1989) suggest that managers make investments that are particularly beneficial to the firm under their own management, thus becoming entrenched.

length of the employment relationship.<sup>8</sup> Furthermore, in general, cash flows depend stochastically on effort and managerial productivity due to noise in performance. Therefore, a turnover policy based solely on observed cash flows cannot induce the optimal sequence of separation decisions. In fact, to sustain the optimal turnover policy, it is essential that the managers keep communicating with the board, e.g., by explaining the determinants of past performances and by describing the firm's prospects under current control. A key role of the optimal contract in our theory is indeed to induce a truthful exchange of information between the managers and the board, in addition to the more familiar role of incentivizing the managers' effort. Unfortunately, this communication is also unlikely to be observable by the econometrician.

On the other hand, if one interprets managerial productivity as coming largely from the managers' own idiosyncratic characteristics, as opposed to the quality of their match with the firm, one should expect such productivity to affect the ease with which a manager finds a new job after separation from his current employer.<sup>9</sup> In this case, our theory predicts that managers with a longer tenure in their previous employment, because they are on average less productive, should experience more difficulty in finding a new job than those with a shorter tenure, a prediction that is consistent with the empirical findings of Fee and Hadlock (2004); the effect of tenure on the base salary in the new job is, however, statistically insignificant in their data set.

We also show that, under the profit-maximizing contract, effort can be optimally incentivized through *linear schemes*. These schemes combine a fixed payment with a bonus which is linear in cash flows.<sup>10</sup> The slope of the linear scheme and hence the induced effort level, increases, on average, with the manager's tenure in the firm. As with retention, this property originates from the assumption that the effect of the managers' initial productivity on their subsequent productivity declines 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 declines over time. As a result, on average, the

---

<sup>8</sup>Note that a model in which effort is exogenously fixed would deliver the potentially misleading prediction of a negative correlation between tenure and performance.

<sup>9</sup>In our model, the managers' outside option is exogenous and constant over time. However, our results extend to a setting in which the managers' outside option depends on their current productivity, provided that such a dependence is not so strong to undermine the possibility of sustaining a truthful revelation of information. Formally, the dependence of the outside option on current productivity must not change the sign of the single-crossing conditions (see footnote 23 for further discussion).

<sup>10</sup>As has been noticed in static settings, linear schemes can often be implemented by offering managers an appropriate combination of fixed pay, bonuses, and stocks/options.

slope of the linear scheme, and hence the sustained effort level, increases with tenure.<sup>11</sup>

**Key normative results.** Turning to the normative results, we find that, when compared to what efficiency requires, the firm’s profit-maximizing contract either induces excessive retention throughout the entire relationship, or *excessive firing at the early stages*, 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.

This result, which follows from endogenizing the principal’s outside option, warns against the prediction that employment relationships tend to become efficient over time — a prediction that is quite pervasive in the dynamic contracting literature with a single agent (see Garrett and Pavan, 2009, for a discussion).

As explained above, from an ex-ante viewpoint (i.e., at the time of hiring), the principal expects to extract all surplus from the relationship with the incumbent after the latter’s tenure grows long enough. As a result, the profit-maximizing effort policy, and the surplus it generates, both converge to their efficient counterparts. 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 expected surplus from starting a new relationship under the efficient contract. The reason is that the new relationship is going to be affected by the same informational and incentive problems as the ones governing the interaction with the incumbent manager. This in turn implies that, from an ex-ante viewpoint, after a sufficiently long tenure, replacement becomes less attractive for a firm maximizing profits than for a planner maximizing efficiency, thus explaining why profit-maximizing firms eventually become excessively lenient when it comes to retention decisions.<sup>12</sup>

This last result suggests that policy interventions aimed at inducing a higher turnover, e.g., by

---

<sup>11</sup>The property that the slope of the incentive scheme increases, on average, with the manager’s tenure in the firm does not originate from the possibility of turnover. It is also a feature of contracts offered in settings with a single manager, where managerial replacement is not an option (see e.g. Garrett and Pavan, 2009). What is specific to this paper is the analysis of the interaction of this property with the dynamics of retention decisions.

<sup>12</sup>This is just a heuristic explanation. The formal argument must take into account the fixed-point nature of the board’s optimization problem.



offering firms a temporary tax reduction after a change in management, or through the introduction of a mandatory retirement age for top employees, can in principle increase welfare.<sup>13</sup> However, such policies might be expected to encounter nontrivial opposition on other grounds.

**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 profit-maximizing contract to the efficient one and establishes the key normative results. All proofs are in the Appendix.

## 1.1 Related literature

Our paper bridges two strands of the literature. On the one hand, recognizing the inherently dynamic nature of firms’ turnover policies, one strand studies turnover in repeated interactions over long horizons. Contrary to the present paper, studies in this literature focus on situations in which the information about managerial productivity (equivalently, the match quality) is symmetric between the principal and the agent.<sup>14</sup> Examples include: (i) models in which productivity is constant but unknown to both the firm and the worker and jointly learned over time, as in Jovanovic (1979, 1984); (ii) models in which productivity is stochastic and persistent but jointly observed as in Acharya (1992) and, in a relational partnership setting, McAdams (2010); and (iii) models in which productivity is constant over time and commonly known. In the latter case, the chief role of termination is to provide incentives for effort to resolve a moral hazard problem with limited liability on the agent’s side (as in the Shapiro and Stiglitz (1984) seminal work on efficiency wages) — more recent examples in this vein include De Marzo and Fishman (2007), De Marzo and Sannikov (2006) and Fong and Li (2010).

What is missing in the works mentioned above is an account of how the managers’ private information affects the dynamics of the relationship — in particular, the equilibrium turnover policy.

---

<sup>13</sup>See, for instance, Lazear (1979) for alternative explanations for why mandatory retirement can be beneficial.

<sup>14</sup>An exception is Yang (2009), who studies an environment in which the agent has private information about his ability to generate output. The agent is either “competent” and, after maximal effort, generates high output with certainty, or is “incompetent” and generates low output with positive probability. Two types of contract are considered that implement high effort by the competent type: “screening contracts” and “pooling contracts”. Under screening contracts, incompetent types reveal themselves in the first period and are terminated. Under pooling contracts, incompetent types reveal themselves only by generating low output, at which point they are fired. The equilibrium turnover policy is therefore a simple consequence of the technology and the two kinds of contracts considered. As a result, the paper’s focus is on characterizing profit-maximizing payment schemes rather than the optimal turnover policy.

The importance of private information has been recognized by a different strand of the literature, which, however, considers only one-shot retention decisions. Eisfeldt and Rampini (2008) and Inderst and Mueller (2008), for example, recognize the importance of incentivizing managers to relinquish the control of the firm’s assets when their privately observed productivity is low.<sup>15</sup> None of these works examine the *dynamics* of turnover policies, which is the focus of the first strand of the literature and of our paper.

The value from bridging these two literatures comes from the fact that it sheds new light on the dynamics of retention decisions. Indeed, allowing managerial productivity to change over time while also assuming that variations in managerial productivity are the managers’ private information is essential for our key positive result about retention policies becoming gradually more lenient over time.<sup>16</sup> Likewise, endogenizing the firm’s outside option by assuming that the relationship with a new manager is affected by the same informational and incentive problems as the one with the incumbent manager is essential for our key normative result about the excessive leniency of firms’ retention policies after long tenure.

Clearly, our work also relates to the literature on managerial compensation which is much too vast to be successfully summarized here. We refer the reader to Garrett and Pavan (2009) for a partial overview.<sup>17</sup> That paper shows how optimal managerial compensation in a dynamic setting can be characterized using a dynamic mechanism design approach that builds on techniques recently developed in Pavan, Segal and Toikka (2009). That paper, however, does not give the firm the possibility of replacing its management, thus abstracting from the dynamic interaction between compensation and retention which is the focus of this paper and is what distinguishes it from the rest of the dynamic mechanism design literature.<sup>18</sup>

---

<sup>15</sup>A related work is Sen (1996), who examines a two-period model where the agent’s only private information is the productivity of the firm itself, independent of the manager who runs it. Commitments to fire the initial manager can be beneficial because they mitigate his information rents.

<sup>16</sup>In an environment where productivity is the managers’ private information but is constant over time, the productivity level required for retention does not vary with the length of the employment relationship (unless one imposes other frictions such as a limited set of feasible contracts). Likewise, in an environment in which productivity changes over time (or is constant but gradually learned by the firm and the manager in a symmetric way, as in Jovanovic 1979), the productivity level required for retention may change over time because it may depend on the amount of available information; however, for given information, it is independent of the number of periods it took for such information to become available. In contrast, in our theory, firms become gradually more lenient over time even when controlling for the amount of information they possess.

<sup>17</sup>The overview in that paper focuses primarily on dynamic models with persistent private information. An excellent overview of the literature on dynamic contracting under moral hazard can be found in Edmans and Gabaix (2009a).

<sup>18</sup>The idea of modelling managerial compensation as a mechanism design problem traces back at least to Laffont and

## 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 agents (the managers). The firm is expected to operate for infinitely many periods and each agent is expected to live as long as the firm. There are infinitely many agents. All agents are ex-ante identical and have a productivity (equivalently, an intrinsic ability to generate cash flows for the firm) that evolves stochastically over time according to the Markov process described below.

**Stochastic process.** Let  $t \in \mathbb{N}$  denote the number of periods that a given agent has been working for the firm. Irrespective of the date of hiring, the agent’s period- $t$  productivity  $\theta_t$  (i.e., his productivity during the  $t^{\text{th}}$  period of employment) is drawn from a cumulative distribution function  $F_t(\cdot|\theta_{t-1})$  defined on the interval  $\Theta_t = [\underline{\theta}_t, \bar{\theta}_t]$  with  $\underline{\theta}_t, \bar{\theta}_t \in \mathbb{R}$ ,  $\bar{\theta}_t > \underline{\theta}_t$ , 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  $\Theta_t$ .<sup>19</sup>

Hereafter, we identify the process governing the evolution of the agents’ productivity with the collection of kernels  $F \equiv \langle F_t(\cdot|\theta_{t-1}) \rangle_{t=1}^{\infty}$ . Let  $\theta^t \equiv (\theta_1, \dots, \theta_t) \in \Theta^t \equiv \times_{s=1}^t \Theta_s$ .<sup>20</sup> For each  $t$ , then let

$$R^t \equiv \{ \theta^t \in \Theta^t : \theta_1 \in \Theta_1 \text{ and } \theta_l \in \text{Supp}[F_l(\cdot|\theta_{l-1})], \text{ all } l = 2, \dots, t \}$$

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  $\theta_{t-1} \in \Theta_{t-1}$ ,  $\partial F_t(\theta_t|\theta_{t-1})/\partial \theta_{t-1}$  exists.<sup>21</sup> Given these assumptions, one can then define the *impulse response functions* associated with

---

Tirole (1986). Garrett and Pavan (2009), as well as Edmans and Gabaix (2009b), and Edmans, Gabaix, Sadzik, and Sannikov (2009) show how this approach can be fruitfully extended to dynamic settings. In common with Garrett and Pavan (2009), the present paper’s chief interest is in characterizing not only the shape of compensation schemes, but also how the dynamics of optimal effort and retention decisions respond to shocks to the environment that are anticipated but, to a large extent, asymmetrically observed. In contrast, the other works assume these dynamics are exogenous and typically focus on the case of constant effort decisions. Another difference is that managers possess private information at the time of contracting so that the contracts the firm offers to its managers must not only incentivize their effort but also screen the managers’ types.

<sup>19</sup> 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 retained with certainty, irrespective of his productivity; or (ii) he is retained 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.

<sup>20</sup> Throughout the entire manuscript, we will use superscripts to denote sequences of variables.

<sup>21</sup> Note that  $\partial F_t(\theta_t|\underline{\theta}_{t-1})/\partial \theta_{t-1}$  (respectively,  $\partial F_t(\theta_t|\bar{\theta}_{t-1})/\partial \theta_{t-1}$ ) denotes the right-hand (respectively, left-hand)

the process  $F$  (see also Pavan, Segal and Toikka, 2009) as follows. For any  $t$  and  $\tau$ ,  $\tau > t$ , and any  $\theta^\tau$ , let  $\theta_{>t}^\tau \equiv (\theta_{t+1}, \dots, \theta_\tau)$ . For all  $t$ , all  $\theta^t \in R^t$ , let  $J_t^t(\theta_t) \equiv 1$ , while for all  $\tau > t$ , all  $\theta^\tau \in R^\tau$ , let

$$J_t^\tau(\theta_{>t-1}^\tau) \equiv \prod_{k=t+1}^\tau I_{k-1}^k(\theta_k, \theta_{k-1}),$$

with each  $I_{k-1}^k$  defined by

$$I_{k-1}^k(\theta_k, \theta_{k-1}) \equiv \frac{-\partial F_k(\theta_k|\theta_{k-1})/\partial \theta_{k-1}}{f_k(\theta_k|\theta_{k-1})}.$$

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

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 agents*.

**Effort, cash flows, and payoffs.** After learning his period- $t$  productivity  $\theta_t$ , the agent 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 agent’s compensation, are given by

$$\pi_t = \theta_t + e_t + \nu_t, \tag{1}$$

where  $\nu_t$  is a transitory noise shock. The shocks  $\nu_t$  are i.i.d. over time, independent across agents, and drawn from the distribution  $\Phi$ , with expectation  $\mathbb{E}[\tilde{\nu}_t] = 0$ . The sequences of productivities  $\theta^t$  and efforts  $e^t \equiv (e_1, \dots, e_t) \in E^t$  are each agent’s private information. In contrast, the history of cash derivative of  $F_t$  with respect to  $\theta_{t-1}$ .

flows  $\pi^t \equiv (\pi_1, \dots, \pi_t) \in \Pi^t \equiv \mathbb{R}^t$  generated by each agent is verifiable and can be used to determine an agent's compensation.

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

$$U^A(c, e) = \sum_{t=1}^{\infty} \delta^{t-1} [c_t - \psi(e_t)], \quad (2)$$

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 agents' compensation. Formally, let  $\pi_{it}$  and  $c_{it}$  denote, respectively, the cash flow generated and the compensation received by the  $i^{\text{th}}$  agent employed by the firm in his  $t^{\text{th}}$  period of employment. Then, let  $T_i$  denote the number of periods for which agent  $i$  works for the firm. The contribution of agent  $i$  to the firm's payoff, evaluated at the time agent  $i$  is hired, is given by

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

Next, denote by  $I \in \mathbb{N} \cup \{+\infty\}$  the total number of agents 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

$$U^P = \sum_{i=1}^I \delta^{\sum_{j=1}^{i-1} T_j} X_i(\pi_i^{T_i}, c_i^{T_i}). \quad (3)$$

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 agents.

**Timing and labor market.** The firm's interaction with the labor market unfolds as follows. In period one (period one is the first period in which the firm uses its new employment contract), the firm is randomly matched with a new manager. Immediately after being matched with the firm, the manager privately learns his period-1 productivity  $\theta_1$  (drawn from the distribution  $F_1$ )<sup>22</sup> and then decides whether or not to sign the employment contract described in full detail below. If the manager

---

<sup>22</sup>The distribution  $F_1$  can also be interpreted as the (unmodelled) steady-state distribution of managerial talent in the market.

refuses to sign, he then leaves the firm, will never be matched with it again, and receives an outside option equal to  $U^o \geq 0$ , irrespective of  $\theta_1$ .<sup>23</sup>

After signing the contract, the agent sends a message  $m_1$  to the firm (think of  $m_1$  either as the revelation of the agent's period-1 type  $\theta_1$  or as the choice of an element of the contract, such as a clause pertaining to the compensation scheme and/or the turnover policy). After sending the message  $m_1$ , the agent 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 then pays the agent compensation  $c_1$  (which may depend on both the clause  $m_1$  and the cash flows  $\pi_1$ ) and the contract stipulates whether or not the agent is to be retained into the second period (in general, we allow retention to depend on  $m_1$  and  $\pi_1$ ).<sup>24</sup>

The agent's second-period productivity is drawn from the distribution  $F_2(\cdot|\theta_1)$ . After privately learning  $\theta_2$ , the agent decides whether or not to leave the firm. If he leaves, the agent 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 agent 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 agent will be retained into the next period.

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

---

<sup>23</sup>In this case, the principal must wait until the next period before being able to go back to the labor market. We assume that the outside option  $U^o$  is sufficiently low that the principal never finds it optimal to have the firm operate without a manager. Of more interest is the assumption that the outside option is independent of type (and time). This assumption seems appropriate when we interpret the agent's period- $t$  type  $\theta_t$  as the quality of his match with the firm, so that his type is firm-specific. It seems less appropriate if one interprets the agent's type as a proxy for managerial talent, which may be correlated across jobs. In this case, it is better to assume that, in each period  $t$ , the agent's outside option is given by a function  $U^o(\theta_t)$ . All our results extend qualitatively to this setting, provided that the derivative of this function is sufficiently small, which is the case, for example, when (i) the discount factor is not very high, and/or (ii) the labor market is "tight" in the sense that an agent who is fired expects to take a long time before finding a new job.

<sup>24</sup>That the contract specifies explicitly a retention policy simplifies the exposition but is not essential. 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 effort and retention decisions by delegating to the agents the choice of whether or not to stay in the relationship.

is randomly matched with a new manager.<sup>25</sup> 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.<sup>26</sup>

**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  $|\theta_t| < K$  for all  $t$  and all  $\theta_t \in \Theta_t$ . We will also assume that the functions  $J_t^\tau(\cdot)$  are uniformly bounded, in the sense that there exists  $\bar{J} < +\infty$  such that  $J_t^\tau(\theta_{>t-1}^\tau) < \bar{J}$  for all  $t$  and  $\tau$ ,  $\tau > t$ , 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  $\psi''(e) > 0$  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 agents are ex-ante identical, time is infinite, and types evolve independently across agents, the firm offers the same contract to each agent. Under such a contract, the compensation the firm pays to each agent is a function of the messages the agent sends and the cash flows that he generates, but

<sup>25</sup>Note that the decision on whether or not to separate from an incumbent manager and go back to the labor market in period  $t + 1$  is determined in period  $t$ . This assumption is meant to capture the idea that it takes time (in the model, one period) to find a replacement. Without such a friction, the board would sample until it found a manager of the highest possible productivity, which would both be unrealistic and make the analysis totally uninteresting.

<sup>26</sup>The assumption of random matching is quite standard in the labor/search literature (see, e.g., Jovanovic, 1979, 1984). In our setting, it implies that whilst the principal's outside option is endogenous, there is no direct competition among workers for employment contracts. This distinguishes the environment from an auction-like setting where, in each period, the principal consults simultaneously with multiple agents and then chooses which one to hire/retain.

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 agents. Hereafter, we will thus maintain the notation that  $t$  denotes the number of periods an agent 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 cash flows are only a noisy transformation of effort and productivity, it is easy to see that under both the efficient and the profit-maximizing contracts these choices can be restricted to depend only on reported productivities  $\theta^t$ . On the other hand, it is essential that the compensation be allowed to depend both on the reported productivities  $\theta^t$  and on past and current cash flows  $\pi^t$ .

A direct revelation mechanism  $\Omega \equiv \langle \xi, s, \kappa \rangle$  comprises sequences of functions  $\xi = (\xi_t : \Theta^t \rightarrow E)_{t=1}^\infty$ ,  $s = (s_t : \Theta^t \times \Pi^t \rightarrow \mathbb{R})_{t=1}^\infty$  and  $\kappa = (\kappa_t : \Theta^t \rightarrow \{0, 1\})_{t=1}^\infty$  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 taken at the end of period  $t$ , with  $\kappa_t(\theta^t) = 1$  if the agent is 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  $\theta_{t+1}$ , and  $\kappa_t(\theta^t) = 0$  if (i) either he is fired at the end of period  $t$ , or (ii) if he was fired in previous periods; i.e.,  $\kappa_t(\theta^t) = 0$  implies  $\kappa_s(\theta^s) = 0$  all  $s > t$ , all  $\theta^s$ .<sup>27</sup> 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 previous reports  $\hat{\theta}^{t-1}$  and cash flows realizations  $\pi^{t-1}$ , the mechanism then operates as follows:

---

<sup>27</sup>Note that, 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.



- At the beginning of period  $t$ , after learning his period- $t$  type  $\theta_t \in \Theta_t$  and upon choosing to stay in the relationship, the agent sends a report  $\hat{\theta}_t \in \Theta_t$ ;
- The mechanism then reacts by prescribing an effort choice  $\xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t)$  and by specifying a reward scheme  $s_t(\hat{\theta}^{t-1}, \hat{\theta}_t, \pi^{t-1}, \cdot) : \Pi_t \rightarrow \mathbb{R}$  and a retention decision  $\kappa_t(\hat{\theta}^{t-1}, \hat{\theta}_t)$ ;
- The agent then chooses effort  $e_t$ ;
- After observing the realized cash flows  $\pi_t = e_t + \theta_t + \nu_t$ , the agent 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  $\kappa_t(\hat{\theta}^{t-1}, \hat{\theta}_t)$ .

By the revelation principle, we restrict attention to direct mechanisms for which (i) a *truthful* and *obedient* strategy is optimal for the agent, and (ii) after any truthful and obedient history, the agent finds it optimal to stay in the relationship whenever offered the possibility of doing so (i.e., the agent 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 as the sum of a representative agent's expected payoff and of the firm's expected profits (the “efficient” policies).<sup>28</sup> Although we are 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 agents' productivities and effort choices are observable and verifiable. In turn, because all players' payoffs are linear in payments, these policies also coincide with the ones that the principal would choose under symmetric information. 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 are implementable also under asymmetric information.

The efficient effort policy is very simple: Because all players are risk neutral and because each agent's productivity has no effect on the marginal cost or the marginal benefit of effort, the efficient

---

<sup>28</sup>Because of symmetry across agents and transferable payoffs, these policies also maximize the ex-ante sum of all agents' payoffs together with the firm's expected profits.

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 agent and continuing experimenting with the incumbent. Define  $A^E \equiv \cup_{t=1}^{\infty}(\Theta_t \times \{t\})$  and denote by  $\mathcal{B}^E$  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  $W^E \in \mathcal{B}^E$  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** *Suppose that the environment satisfies the conditions of the model setup. The efficient effort and turnover policies satisfy the following properties.<sup>29</sup> (i) For all  $t$ , all  $\theta^t \in R^t$ ,  $\xi_t^E(\theta^t) = e^E$ , with  $e^E$  implicitly defined by*

$$\psi'(e^E) = 1. \quad (4)$$

(ii) *There exists a sequence of thresholds  $(\theta_t^E)_{t=1}^{\infty}$  such that, conditional on being employed in period  $t$ , the agent is retained at the end of period  $t$  if and only if  $\theta_t \geq \theta_t^E$ .<sup>30</sup> The thresholds  $(\theta_t^E)_{t=1}^{\infty}$  are defined as follows. Let  $W^E$  be the unique fixed point to the mapping  $T_E : \mathcal{B}^E \rightarrow \mathcal{B}^E$  defined, for all  $W \in \mathcal{B}^E$ , all  $(\theta_t, t) \in A^E$ , by*

$$\begin{aligned} T_E W(\theta_t, t) &= \theta_t + e^E - \psi(e^E) - (1 - \delta)U^o \\ &+ \delta \max \left\{ \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W(\tilde{\theta}_{t+1}, t+1) \right]; \mathbb{E}_{\tilde{\theta}_1} \left[ W(\tilde{\theta}_1, 1) \right] \right\}. \end{aligned} \quad (5)$$

For any  $t$ ,

$$\theta_t^E = \inf \{ \theta_t \in \Theta_t : \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W^E(\tilde{\theta}_{t+1}, t+1) \right] \geq \mathbb{E}_{\tilde{\theta}_1} \left[ W^E(\tilde{\theta}_1, 1) \right] \},$$

unless the set is empty, in which case  $\theta_t^E = \bar{\theta}_t$ .

<sup>29</sup>The efficient policies are "essentially unique", i.e., unique up to a zero-measure set of histories.

<sup>30</sup>Again, it should be clear that it is only for expositional convenience that we allow the effort and retention policies to be defined in each period for all possible histories, including those that lead to separation at earlier periods.

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  $T_E : \mathcal{B}^E \rightarrow \mathcal{B}^E$  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 fact that the process is Markov and satisfies the property of first-order stochastic dominance in types, in turn establish that the efficient turnover policy is the cut-off rule given in the proposition.

## 4 The profit-maximizing contract

We now turn to the characterization of the firm's profit-maximizing contract in a setting where neither the agents' productivities nor their effort choices are observable. We begin by providing some sufficient conditions for an effort policy  $\xi$  and a turnover policy  $\kappa$  to be implementable.<sup>31</sup> Note that these policies are defined over the supports of the marginal distributions and thus specify effort and retention decisions also for off-support histories  $\theta^t \notin R^t$ . The reason for extending the policies from  $R^t$  to  $\Theta^t$  is that this permits one 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 characterization of incentive compatibility on the equilibrium path.

**Proposition 2** *Suppose that the conditions of the model setup hold. Let  $\xi$  and  $\kappa$  be effort and turnover policies such that each  $\xi_t$  and  $\kappa_t$  are independent of  $\pi^{t-1}$  (respectively  $\pi^t$ ), all  $t$ . In addition, suppose that  $\xi$  and  $\kappa$  are such that the following single-crossing conditions hold for all  $t$ , all  $\hat{\theta}^{t-1} \in \Theta^{t-1}$  such that  $\kappa_{t-1}(\hat{\theta}^{t-1}) = 1$ , all  $\theta_t, \hat{\theta}_t \in \Theta_t$ :*

$$\mathbb{E}_{\hat{\theta}_{>t}^\infty | \theta_t} \left[ \begin{array}{l} \sum_{k=t}^{\tau(\hat{\theta}^{t-1}, \theta_t, \hat{\theta}_{>t}^\infty)} \delta^{k-t} J_t^k(\theta_t, \tilde{\theta}_{>t}^k) \psi' \left( \xi_k(\hat{\theta}^{t-1}, \theta_t, \tilde{\theta}_{>t}^k) \right) \\ - \sum_{k=t}^{\tau(\hat{\theta}^{t-1}, \hat{\theta}_t, \hat{\theta}_{>t}^\infty)} \delta^{k-t} J_t^k(\theta_t, \tilde{\theta}_{>t}^k) \psi' \left( \xi_k(\hat{\theta}^{t-1}, \hat{\theta}_t, \tilde{\theta}_{>t}^k) \right) \end{array} \right] [\theta_t - \hat{\theta}_t] \geq 0. \quad (6)$$

Then there exists a **linear reward scheme** of the form

$$s_t(\theta^t, \pi_t) = S_t(\theta^t) + \alpha_t(\theta^t)\pi_t \text{ all } t, \text{ all } \theta^t \in \Theta^t,$$

such that, irrespective of the distribution  $\Phi$  of the transitory noise, the following are true: (i) in each period  $t \geq 1$ , the agent finds it optimal to follow a truthful and obedient strategy, irrespective of the

<sup>31</sup>By “implementable” we mean that there exists a compensation scheme  $s$  that, given  $\xi$  and  $\kappa$ , induces the agent to follow a truthful and obedient strategy.

history of past reports  $\hat{\theta}^{t-1}$ , effort choices  $e^{t-1}$ , and cash flow realizations  $\pi^{t-1}$ ; (ii) by following a truthful and obedient strategy, the lowest period-1 type  $\underline{\theta}_1$  expects a payoff equal to his outside option; and (iii) at the beginning of any period, the agent's continuation payoff from following a truthful and obedient strategy in the current and future periods is at least as high as his outside option.

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. A special case of interest is when both the turnover policy  $\kappa$  and the effort policy  $\xi$  are *strongly monotone*, i.e., when each  $\xi_t(\cdot)$  and  $\kappa_t(\cdot)$  is nondecreasing. For example, this property is satisfied by the efficient policies characterized in the previous section, 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 agents' productivities) that guarantee that the property is satisfied also by the policies that maximize the firm's expected profits. As the result in the proposition makes clear, these conditions are stronger than necessary. In fact, linear schemes permit the firm to implement also 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 firm, or by the agents, 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 agent 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 schemes, the coefficient

$$\alpha_t(\theta^t) \equiv \psi'(\xi_t(\theta^t)) \tag{7}$$

is chosen so as to provide the agent 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 current pay to the current cash flows is given by (7), by choosing effort  $e_t = \xi_t(\theta^t)$ , the agent equates the marginal disutility of effort to its marginal benefit and hence maximizes his continuation payoff. This is irrespective of whether or not the agent has reported his productivity truthfully. Once effort is controlled by the variable component  $\alpha_t(\theta^t)$ , the fixed component  $S_t(\theta^t)$  can be chosen so as to guarantee that the agent has the right incentives to report his productivity truthfully. As we show in the Appendix, this is accomplished by setting

$$\begin{aligned}
S_t(\theta^t) \equiv & \psi(\xi_t(\theta^t)) - \alpha_t(\theta^t) (\xi_t(\theta^t) + \theta_t) + (1 - \delta) U^o \\
& + \int_{\underline{\theta}_t}^{\theta_t} \mathbb{E}_{\tilde{\theta}_{>t}^\infty | s} \left[ \sum_{k=t}^{\tau(\theta^{t-1}, s, \tilde{\theta}_{>t}^\infty)} \delta^{k-t} J_t^k(s, \tilde{\theta}_{>t}^k) \psi'(\xi_k(\theta^{t-1}, s, \tilde{\theta}_{>t}^k)) \right] ds \\
& - \delta \kappa_t(\theta^t) \mathbb{E}_{\tilde{\theta}_{t+1} | \theta_t} [u_{t+1}(\tilde{\theta}_{t+1}; \theta^t)]
\end{aligned} \tag{8}$$

where

$$u_{t+1}(\theta_{t+1}; \theta^t) \equiv \int_{\underline{\theta}_{t+1}}^{\theta_{t+1}} \mathbb{E}_{\tilde{\theta}_{>t+1}^\infty | s} \left[ \sum_{k=t+1}^{\tau(\theta^t, s, \tilde{\theta}_{>t+1}^\infty)} \delta^{k-(t+1)} J_{t+1}^k(s, \tilde{\theta}_{>t+1}^k) \psi'(\xi_k(\theta^t, s, \tilde{\theta}_{>t+1}^k)) \right] ds \tag{9}$$

denotes the agent's period-(t+1) continuation payoff (over and above his outside option) under the truthful and obedient strategy.<sup>32</sup> Combining the two components  $\alpha$  and  $S$  one can then verify that, when the policies  $\xi$  and  $\kappa$  satisfy the single-crossing conditions in the proposition, then after *any* history  $(\theta^t, \hat{\theta}^{t-1}, e^{t-1}, \pi^{t-1})$ , the agent 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.

We have shown that linear compensation schemes permit the principal to sustain a fairly rich set of effort and turnover policies. By verifying the single-crossing conditions of Proposition 2, we now show

<sup>32</sup>By inspecting the formula for the fixed component  $S_t(\theta^t)$  in (8), one can verify that when either (i) the support of the transitory noise distribution  $\Phi$  is bounded and the level of the outside option  $\underline{U}$  is sufficiently large, or (ii) the infimum of all possible noise realizations  $\inf\{Supp[\Phi]\}$  and the discount factor  $\delta$  are close to zero, then the linear reward schemes of Proposition 2 entail a nonnegative payment to the agent in every period and for every profit realization consistent with a truthful and obedient strategy. In these cases, the proposed schemes offer a valid implementation also in settings in which the managers are protected by limited liability.

that, under certain regularity conditions, the effort and turnover policies that maximize the firm's expected profits indeed belong to this set

From arguments similar to those in Pavan, Segal, and Toikka (2009),<sup>33</sup> one can verify that, in *any* incentive-compatible mechanism  $\Omega \equiv \langle \xi, s, \kappa \rangle$ , each manager's period-1 expected payoff under a truthful and obedient strategy  $V^\Omega(\theta_1)$  must satisfy

$$V^\Omega(\theta_1) = V^\Omega(\underline{\theta}_1) + \int_{\underline{\theta}_1}^{\theta_1} \mathbb{E}_{\tilde{\theta}_{>1}^\infty | s} \left[ \sum_{t=1}^{\tau(s, \tilde{\theta}_{>1}^\infty)} \delta^{t-1} J_1^t(s, \tilde{\theta}_{>1}^t) \psi'(\xi_t(s, \tilde{\theta}_{>1}^t)) \right] ds. \quad (10)$$

Importantly, note that the formula in (10) must hold irrespective of whether or not the compensation is linear. That (10) is necessary for incentive-compatibility in fact follows by applying envelope arguments similar to those for static optimization, adjusted to account for the dynamics of the productivity process.

The formula in (10) confirms the intuition that the surplus that the principal must leave to each type to induce him to reveal himself is determined by the dynamics of *effort* and *retention* decisions under the contracts offered to 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, since they are able to generate the same distribution of cash flows by working less. The amount of effort they expect to save by mimicking must, however, take into account the fact that both their own productivity as well as that of the types they mimic will change over time. This is done by weighting the amount of effort saved in all subsequent periods by the impulse response functions  $J_1^t$ , which, as explained above, control for how the effect of the initial productivity on future productivity evolves over time.

Using (10), the firm's expected profits from hiring each manager can be conveniently expressed as

$$\mathbb{E}_{\tilde{\theta}^\infty, \tilde{\nu}^\infty} \left[ \sum_{t=1}^{\tau(\tilde{\theta}^\infty)} \delta^{t-1} \left\{ \begin{array}{l} \tilde{\theta}_t + \xi_t(\tilde{\theta}^t) + \tilde{\nu}_t - \psi(\xi_t(\tilde{\theta}^t)) \\ -\eta(\tilde{\theta}_1) J_1^t(\tilde{\theta}^t) \psi'(\xi_t(\tilde{\theta}^t)) - (1-\delta) U^o \end{array} \right\} \right] + U^o - V^\Omega(\underline{\theta}_1), \quad (11)$$

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  $F_1$ . The formula in (11) is the dynamic analog of the familiar *virtual surplus* formula for static adverse selection settings.

<sup>33</sup>See also Garrett and Pavan (2009) for an illustration of how these arguments must be adapted to a moral hazard setting similar to the one in the current paper.

It expresses the firm's profits as the discounted expected total surplus of the relationship net of terms that control for the surplus that the firm must leave to the managers (over and above their outside option) to induce them to truthfully reveal their private information.

Equipped with the aforementioned representation, now consider the “*relaxed program*” that consists of choosing policies  $(\xi_t(\cdot), \kappa_t(\cdot))_{t=1}^\infty$  so as to maximize the sum of the profits generated by each manager, as given by (11), subject to the participation constraint for the lowest period-1 type:  $V^\Omega(\theta_1) \geq U^o$ . Hereafter, we denote by  $(\xi_t^*(\cdot), \kappa_t^*(\cdot))_{t=1}^\infty$  the policies that maximize (11). Using the result in Proposition 2, it is then easy to see that these policies are indeed the ones that maximize the firm's expected profits, provided that there exists an extension of these policies from  $R$  to  $\Theta$  such that the extended policies satisfy all the single-crossing conditions. In the following proposition, we first characterize the policies  $(\xi_t^*(\cdot), \kappa_t^*(\cdot))_{t=1}^\infty$  that maximize (11). We then provide a simple sufficient condition for the existence of an extension from  $R$  to  $\Theta$ . (Recall that the role of these extensions is to permit the agents to truthfully reveal their types also after histories that involve a departure from truthful and obedient behavior in past periods.)

Let  $A \equiv \cup_{t=1}^\infty (R^t \times \{t\})$  and denote by  $\mathcal{B}$  the space of bounded functions from  $A$  to  $\mathbb{R}$ .

**Proposition 3** *Suppose that the conditions of the model setup hold. Consider the policies  $\xi^*$  and  $\kappa^*$  defined by (i) and (ii) below.<sup>34</sup> (i) For all  $t$ , all  $\theta^t \in R^t$ , the effort policy  $\xi_t^*(\theta^t)$  is implicitly given by<sup>35</sup>*

$$\psi'(\xi_t^*(\theta^t)) = 1 - \eta(\theta_1) J_1^t(\theta^t) \psi''(\xi_t^*(\theta^t)). \quad (12)$$

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

by

$$TW(\theta^t, t) \equiv \xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1) J_1^t(\theta^t) \psi'(\xi_t^*(\theta^t)) - (1 - \delta)U^o + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t} \left[ W(\tilde{\theta}^{t+1}, t+1) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W(\tilde{\theta}_1, 1) \right] \right\}. \quad (13)$$

For all  $t$ , all  $\theta^t \in R^t$ , the retention policy  $\kappa^*$  is such that, conditional on being employed in period  $t$ , the manager is retained at the end of period  $t$  if and only if

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

<sup>34</sup>Once again, it is only for expositional convenience that we allow the effort and retention policies to be defined in each period for all possible histories, including those that lead to separation at earlier periods.

<sup>35</sup>For simplicity, we assume throughout that the profit-maximizing policy specifies positive effort choices in each period  $t$  and for each history  $\theta^t \in R^t$ . This amounts to assuming that, for all  $t$  all  $\theta^t \in R^t$ ,  $\psi''(0) < 1/[\eta(\theta_1) J_1^t(\theta^t)]$ . When this condition does not hold, optimal effort is simply given by  $\xi_t^*(\theta^t) = 0$ .

Suppose that there exists an extension of the policies  $(\xi^*, \kappa^*)$  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  $(\xi^*, \kappa^*)$  given above. A sufficient condition for such an extension to exist is that each function  $\eta(\cdot)J_1^t(\cdot)$  is nonincreasing on  $R^t$ , 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  $(\theta_t^*(\cdot))_{t=1}^\infty$ ,  $\theta_t^* : R^{t-1} \rightarrow \mathbb{R}$ , all  $t \geq 1$ ,<sup>36</sup> such that, conditional on being employed in period  $t$ , the manager is retained at the end of period  $t$  if and only if  $\theta_t \geq \theta_t^*(\theta^{t-1})$ , with  $\theta_t^*(\cdot)$  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 the effort policy  $\xi^*$  is obtained by maximizing virtual surplus period by period and state by state. In particular, in each period  $t$ , and for each history  $\theta^t \in R^t$ , the optimal effort  $\xi_t^*(\theta^t)$  is chosen so as to trade off the effect of a marginal variation in effort on total surplus  $e_t + \theta_t - \psi(e_t) - (1 - \delta)U^o$  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. As a result, the policy  $\xi^*$  that maximizes "virtual surplus" (11) is independent of the turnover policy  $\kappa$ . 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  $\xi_t(\hat{\theta}_1, \theta_{>1}^t)$  that the firm asks each less productive type  $\hat{\theta}_1 < \theta_1$  to exert in each period  $t \geq 1$ , the firm's profit-maximizing effort policy 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  $J_1^t$ . These functions, by describing the effect of period-one productivity on subsequent productivity, capture how the persistence of the agents' 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

<sup>36</sup>Because  $\theta_0$  is given, in period one, the cut-off function  $\theta_1^* : R^0 \rightarrow \mathbb{R}$  reduces to a single cut-off  $\theta_1^* \in \mathbb{R}$ .



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. 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 agent by (11)) 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, taking into account its effect on the managers' informational rents, as expected at the time they are hired. As shown above, these rents are determined by the dynamics of the impulse response functions  $J_1^t$ . Because the latter typically depend on the entire history of productivity realizations  $\theta^t$ , so does the profit-maximizing turnover policy  $\kappa^*$ .<sup>37</sup>

The profit-maximizing turnover policy can then be determined straightforwardly from  $W^*$ : the incumbent manager is fired whenever the expected value  $\mathbb{E}_{\tilde{\theta}_1} \left[ W^*(\tilde{\theta}_1, 1) \right]$  of starting a relationship with a new manager of unknown productivity exceeds the expected value  $\mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^*(\tilde{\theta}^{t+1}, t + 1) \right]$  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.

---

<sup>37</sup>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. This can be seen by considering the flow virtual surplus defined in (14) below.

Having characterized the policies that maximize the firm’s virtual surplus (11), 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,<sup>38</sup> the profits

$$VS_t(\theta^t) \equiv \xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1)J_1^t(\theta^t)\psi'(\xi_t^*(\theta^t)) - (1 - \delta)U^o \quad (14)$$

that the firm expects in each period  $t$  from an incumbent manager (net of information rents) are nondecreasing in  $\theta^t$ . 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 firm 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 type  $\theta_1$  an expected payoff equal to his outside option.<sup>39</sup> 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 (11) under any mechanism that is incentive compatible and individually rational for the manager.

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  $J_t^\tau(\cdot)$  are uniformly bounded over  $t, \tau, \tau > t$ , and  $R^\tau$ ; (iv) each function  $\eta(\cdot)J_1^t(\cdot)$  is nonincreasing. Although these

<sup>38</sup>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  $J_1^t$ .

<sup>39</sup>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.

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.

**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  $\sum_{s=1}^t (\times_{\tau=s+1}^t \gamma_\tau)$  is bounded uniformly across  $t$ .<sup>40</sup> For each  $t$ , let  $G_t$  be an absolutely continuous c.d.f. with support  $[\underline{\varepsilon}_t, \bar{\varepsilon}_t]$ , where  $|\underline{\varepsilon}_t|$  and  $|\bar{\varepsilon}_t|$  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  $t \geq 2$ ,  $\theta_t = \gamma_t \theta_{t-1} + \varepsilon_t$ , with  $\varepsilon_t$  drawn from  $G_t$ , independently from  $\theta_1$  and  $\varepsilon_{-t}$ . 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  $\Theta' = [\underline{\theta}, \bar{\theta}] \supset \Theta_1$  and take an arbitrary continuously differentiable function  $\varphi : \Theta' \rightarrow \mathbb{R}_{++}$  satisfying (i)  $\underline{\theta} \leq \bar{\theta} - \varphi(\underline{\theta})$ , (ii)  $\varphi'(\cdot)/\varphi(\cdot)$  nondecreasing, and (iii)  $\varphi'(\theta) \in [-1, 0]$  for all  $\theta \in \Theta'$ .<sup>41</sup> 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  $t \geq 2$ ,  $\theta_t = \bar{\theta} - \varphi(\theta_{t-1}) \varepsilon_t$ , with  $\varepsilon_t$  drawn from  $G_t$ , independently from  $\theta_1$  and  $\varepsilon_{-t}$ . 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  $(\theta^t, \theta_{>t}^s)$ ,  $\theta_s \geq \theta_t$  implies that  $J_1^s(\theta^t, \theta_{>t}^s) \leq J_1^t(\theta^t)$ .

<sup>40</sup> An example is an autonomous AR(1) process with  $\gamma_t = \gamma \in (0, 1)$  all  $t$ .

<sup>41</sup> An example is  $\varphi(\theta) = e^{-\theta}$  for all  $\theta \in [\underline{\theta}, \bar{\theta}]$ , with  $\underline{\theta} > 0$  and  $\bar{\theta} - e^{-\bar{\theta}} \geq \underline{\theta}$ .

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  $\theta^s = (\theta^t, \theta_{>t}^s) \in R^s$ . Then  $\theta_s \geq \theta_t$  implies that*

$$\mathbb{E}_{\tilde{\theta}^{s+1}|\theta^s} \left[ W^* \left( \tilde{\theta}^{s+1}, s+1 \right) \right] \geq \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right]. \quad (15)$$

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 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 (14), 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  $\theta^s \in R^s$  be such that  $\kappa_{s-1}^*(\theta^{s-1}) = 1$ . 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 agent 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 firing).

The result that the optimal turnover policy becomes more permissive over time together with the result that the productivity level  $\theta_t^*(\theta^{t-1})$  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. However, it is important to recognize that, while this property holds for given sequences of productivities, it need not hold when averaging across the entire pool of productivities of retained managers since composition effects can push in the opposite direction. 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. It is thus essential for an econometrician testing for our key positive prediction to collect data that either directly or indirectly reveal the evolution of individual productivities.

## 5 Efficient vs profit-maximizing turnover

Having established the key positive result that retention becomes gradually more permissive over time, we now turn to the normative implications of this property. We start by verifying that the efficient effort and turnover policies characterized 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 efficient effort and turnover policies of Proposition 1.*<sup>42</sup>

---

<sup>42</sup>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.

We can now compare the firm's 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 gradually increases with time and converges to its efficient level as the length of the employment relationship grows sufficiently large. One might therefore expect the profit-maximizing contract to become gradually closer to its efficient counterpart over time.

This logic, however, fails to take into account that the principal's outside option is endogenous and is shaped by the same information and incentive problems that govern the relationship with the incumbent manager. Taking this endogeneity into account leads to our key normative result that, once the length of the employment relationship has grown sufficiently large, the firm's optimal turnover policy eventually becomes excessively permissive as compared to what efficiency requires. We formalize this result in Proposition 5 below. Before doing that, as a first step to understanding the result, we first consider the following simplified example.

**Example 3** *Suppose that the firm operates for only two periods and that this is commonly known. In addition, suppose that both  $\theta_1$  and  $\varepsilon_2$  are uniformly distributed over  $[-\frac{1}{2}, +\frac{1}{2}]$  and that  $\theta_2 = \gamma\theta_1 + \varepsilon_2$ . Finally, suppose that  $\psi(e) = e^2/2$  for all  $e \in [0, 1]$ , 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  $\gamma < 0.845$ ).*

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  $\theta_1^E = 0$ ).

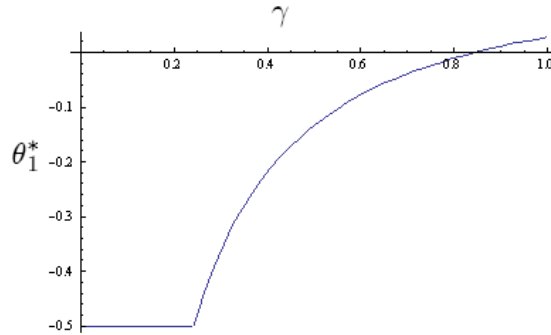


Figure 1: Optimal thresholds  $\theta_1^*$  for retention for different impulse responses  $J_1^2 = \gamma$ .

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,  $VS_1(\theta_1) = \mathbb{E}_{\tilde{\theta}_2|\theta_1} [VS_2(\theta_1, \tilde{\theta}_2)]$ , where the functions  $VS_1$  and  $VS_2$  are given by (14). In this example,  $VS_1$  is strictly convex. Noting that  $\theta_1^E = \mathbb{E}[\tilde{\theta}_1]$ , we then have that  $\mathbb{E}[VS_1(\tilde{\theta}_1)] > VS_1(\theta_1^E) = \mathbb{E}_{\tilde{\theta}_2|\theta_1^E} [VS_2(\theta_1^E, \tilde{\theta}_2)]$ , i.e., the expected value of replacing the agent is greater than the value from keeping the agent 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.

The above comparative statics exercise has a natural counterpart in a dynamic setting with more than two periods. Provided that the impulse responses with respect to the first-period productivity eventually become small, the effect of tenure on retention is similar to that of a reduction in  $\gamma$  in the two-period example above. We start by showing that, when the impulse responses with respect to first-period productivity vanish uniformly over histories once the length of the employment relationship has grown sufficiently large, then, in the long run, the firm's optimal contract always results in inefficiently low turnover (i.e., excessively high retention).

**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  $t > t_\epsilon$ ,  $\eta(\theta_1) J_1^t(\theta^t) < \epsilon$  for all  $\theta^t \in R^t$ .

Under this condition, which again seems plausible for most cases of interest, we 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  $\theta_t \geq \theta_t^E$ ,*

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

(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  $\theta_1 \in \Theta_1$ , the function  $\eta(\theta_1) J_1^t(\theta_1, \cdot)$  is Lipschitz continuous over  $\Theta^t(\theta_1) \equiv \{\theta_{>1}^t \in \Theta_{>1}^t : (\theta_1, \theta_{>1}^t) \in R^t\}$  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  $\kappa_{t-1}^*(\theta^{t-1}) = 1$ , if  $\theta_t^E \in \text{int}\{ \text{Supp}[F_t(\cdot|\theta_{t-1})] \}$ , then  $\theta_t^*(\theta^{t-1}) < \theta_t^E$ .*

Part (i) of Proposition 5 establishes that, if the agent is retained in period  $t > \bar{t}$  under the efficient policy when his productivity is  $\theta_t$ , then he is also retained under the profit-maximizing policy when his period- $t$  productivity is the same. The additional conditions of Part (ii) imply continuity in  $\theta_t$  of the expected continuation payoffs  $\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{t-1}, \cdot)} \left[ W^* \left( \theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1 \right) \right]$  and  $\mathbb{E}_{\tilde{\theta}_{t+1}^E} \left[ W^E \left( \tilde{\theta}_{t+1}, t+1 \right) \right]$  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 managers. In this environment, the firm optimally dictates to 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 other manager.



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,<sup>43</sup> 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  $\underline{t}, \bar{t} \in \mathbb{N}$ , with  $2 \leq \underline{t} \leq \bar{t}$ , such that the following are true: (a) for any  $t < \underline{t}$ , and almost any  $\theta^t \in R^t$ ,*

---

<sup>43</sup>Recall that welfare under the efficient contract with asymmetric information coincides with the sum of the firm’s expected profit and of the agents’ outside option under the contract that the firm would offer if information about all managers’ effort and productivities were symmetric.

if  $\kappa_{t-1}^E(\theta^{t-1}) = 1$  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  $\kappa_{t-1}^*(\theta^{t-1}) = 1$  and  $\kappa_t^*(\theta^t) = 0$ , then  $\kappa_t^E(\theta^t) = 0$ .

The firm's profit-maximizing contract thus either induces excessive retention (i.e., too little turnover) throughout the entire relationship, or it induces excessive firing at the early stages followed by excessive retention in the long run. Formally, there exist dates  $\underline{t}, \bar{t}$  such that any manager who is fired at the end of period  $t < \underline{t}$  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.<sup>44</sup>

As anticipated in the Introduction, these last results suggest that policies aimed at inducing firms to grant their top employees more job stability at the early stages of the employment relationship and less stability at the later stages (e.g., through tax incentives or through the introduction of explicit mandatory retirement ages) can have a positive effect on welfare. Clearly, such policies may have limited appeal (or be infeasible) for reasons beyond the scope of this article.

## 6 Conclusions

The purpose of the paper was to provide a tractable, yet rich, model of managerial contracting that can be used to shed light on properties of the employment relationship that originate from the interaction between compensation and retention in dynamic environments.

The model explicitly accounts for the following possibilities: (i) turnover is driven by variations in the managers' ability to generate profits for the firm; (ii) shocks to managerial productivity are anticipated but privately observed by the managers; (iii) the firm can always go back to the labor market and replace an incumbent manager with a new one—however, the firm's prospects under the new management are bound to be affected by similar incentives and informational problems as the ones governing the relationship with the incumbent management.

Allowing for the aforementioned possibilities not only adds realism to the model but sheds light on certain properties of the employment relationship. In particular, a key role of the contract in a

---

<sup>44</sup>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.

dynamic environment with asymmetric information is to induce managers to explain to the board the determinants of the firm's past performance and future prospects, over and above the more familiar role of incentivizing their effort. A continuous exchange of information between the board and the management is in fact essential to guarantee a prompt response to variations in the environment that may call for managerial turnover.

On the positive side, we showed how the firm's desire to retain flexibility while limiting the level of managerial compensation calls for job instability early on in the relationship followed by job security and entrenchment after sufficiently long tenure. On the normative side, we showed that, compared to what is efficient, the contract that the firm offers to its top employees induces either excessive retention throughout the entire relationship (i.e., insufficiently low turnover), or excessive firing at the early stages followed by excessive retention in the long run.

Throughout the entire analysis, we maintained the assumption of a single firm interacting with multiple managers. In particular, we took as given the managers' outside option and instead focused on endogenizing the firm's one. We also assumed that the matching between the firm and its top employees is governed by an exogenous matching function. These assumptions were instrumental to retain tractability while shedding light on the properties described above. Extending the analysis to richer environments with competing firms would permit one to endogenize also the managers' outside options and thereby the entire matching process. This is an important, yet challenging, direction for future research which is likely to shed further light on the interaction between compensation and retention in dynamic environments.

## References

- [1] Acharya, Sankarshan (1992). 'Maximizing the Market Value of a Firm to Choose Dynamic Policies for Managerial Hiring', *International Economic Review*, 33, 373-397.
- [2] Caillaud Bernard, Roger Guesnerie and Patrick Rey (1992). 'Noisy Observation in Adverse Selection Models', *Review of Economic Studies*, 59, 595-615.

- [3] DeMarzo, Peter and Michael Fishman (2007), ‘Optimal Long-Term Financial Contracting’, *Review of Financial Studies*, 20, 2079-2127.
- [4] DeMarzo, Peter and Yuliy Sannikov (2006), ‘Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model’, *Journal of Finance*, 61, 2681-2724.
- [5] Edmans, Alex and Xavier Gabaix (2009a) ‘Tractability in Incentive Contracting,’ mimeo NYU Stern and Wharton.
- [6] Edmans, Alex and Xavier Gabaix (2009b) ‘Is CEO Pay Really Inefficient? A Survey of New Optimal Contracting Theories,’ *European Financial Management*, 15(3), 486-496.
- [7] Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov (2009), ‘Dynamic Incentive Accounts,’ mimeo NYU Stern, Princeton and Wharton.
- [8] Eisfeldt, Andrea L. and Adriano A. Rampini (2008) ‘Managerial incentives, capital reallocation, and the business cycle’, 87, 177-199.
- [9] Fee, Edward and Charles Hadlock (2004), ‘Management turnover across the corporate hierarchy’, *Journal of Accounting and Economics*, 37, 3-38.
- [10] Fine, Charles H. (1998), *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*, Perseus Books, Cambridge, MA.
- [11] Fong, Yuk-fai and Jin Li (2010), ‘Relational Contracts, Limited Liability, and Employment Dynamics’, mimeo Northwestern University.
- [12] Fudenberg, Drew and David Levine (1983), ‘Subgame-Perfect Equilibria of Finite and Infinite Horizon Games’, *Journal of Economic Theory*, 31, 251-268.
- [13] Garrett, Daniel and Alessandro Pavan (2009) ‘Dynamic Managerial Compensation: A Mechanism Design Approach’, mimeo Northwestern University.
- [14] Inderst, Roman and Holger Mueller (2008), ‘CEO Replacement under Private Information’, mimeo University of Frankfurt and New York University.

- [15] Jovanovic, Boyan (1979), 'Job Matching and the Theory of Turnover', *Journal of Political Economy*, 87, 972-990.
- [16] Jovanovic, Boyan (1984), 'Matching, Turnover, and Unemployment', *Journal of Political Economy*, 92, 108-122.
- [17] Laffont, Jean-Jacques and Jean Tirole (1986), 'Using cost observation to regulate firms', *Journal of Political Economy*, 94, 614-641.
- [18] Lazear, Edward P. (1979), 'Why is There Mandatory Retirement', *Journal of Political Economy*, 87, 1261-1284.
- [19] McAdams, David (2010), 'Performance Turnover in a Stochastic Partnership', mimeo Duke University.
- [20] Miller, Danny (1991), 'Stale in the Saddle: CEO Tenure and the Match between Organization and Environment', *Management Science*, 37, 34-52.
- [21] Pavan, Alessandro, Ilya Segal and Juuso Toikka (2009), 'Dynamic Mechanism Design: Incentive Compatibility, Profit Maximization and Information Disclosure', mimeo Northwestern University and Stanford University.
- [22] Rose, Nancy L. and Andres Shepard (1997), 'Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?' *RAND Journal of Economics*, 28, 489-514.
- [23] Sen, Arijit (1996), 'Termination Clauses in Long-Term Contracts', *Journal of Economics and Management Strategy*, 5, 473-496.
- [24] Shapiro, Carl and Joseph Stiglitz (1984), 'Equilibrium Unemployment as a Worker Discipline Device', *American Economic Review*, 74, 433-444.
- [25] Shleifer, Andrei and Robert W. Vishny (1989), 'Management Entrenchment: the Case of Manager-Specific Investments', *Journal of Financial Economics*, 25, 123-139.

[26] Stokey, Nancy, Peter Lucas with Edward Prescott (1989), *Recursive Methods in Economic Dynamics*. Cambridge, MA: Harvard University Press.

[27] Yang, Huanxing (2009), ‘Nonstationary Relational Contracts’, mimeo Ohio State University.

## Appendix

**Proof of Proposition 1.** That the efficient effort policy is given by  $\xi_t^E(\theta^t) = e^E$  for all  $t$ , all  $\theta^t$ , follows directly from inspection of the principal’s payoff (3), the agents’ payoff (2), and the definition of cash flows given by (1) in the main text.

Consider the retention policy. Because all agents are ex-ante identical, and because the process governing the evolution of the agents’ productivities is Markov, it is immediate that, in each period, the decision of whether or not to retain an agent must depend only on the length of the employment relationship  $t$  and on the agent’s current productivity  $\theta_t$ . We will denote by  $W^E : A^E \rightarrow \mathbb{R}$  the value function associated with the problem that consists of 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 current agent’s productivity is  $\theta_t$  and the agent has been working for the firm already for  $t - 1$  periods (the assumption that the firm must wait at least one period before being matched with a new agent, together with the assumption that each agent’s outside option  $U^o$  is not too large (equivalently, that the surplus lost from operating without a manager is large), implies that the earliest the firm can efficiently replace its current manager is period  $t + 1$ ).

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. We now establish the existence and uniqueness of such a fixed point and show that it is nondecreasing in the productivity level. 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  $\theta_t$ . 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,  $\mathcal{N}^E$  is closed under  $T_E$ . Moreover, “Blackwell’s sufficient conditions” (namely “monotonicity” and “discounting”, the latter being 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  $W \in \mathcal{N}^E$ ,  $\hat{W}^E = \lim_{n \rightarrow \infty} T_E^n W$  exists, is unique, and belongs to  $\mathcal{N}^E$ .

Finally, we conjecture 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  $\mathbb{E} \left[ \hat{W}^E \left( \hat{\theta}_{t+1}, t+1 \right) \mid \theta_t \right] \geq \mathbb{E}_{\hat{\theta}_1} \left[ \hat{W}^E \left( \hat{\theta}_1, 1 \right) \right]$  and  $\kappa_t(\theta^t) = 0$  otherwise. Note that, because the process satisfies the assumption of first-order-stochastic dominance in types, and because  $\hat{W}^E(\cdot, t)$  is nondecreasing for each  $t$ , this retention policy is a cut-off policy. This, together with the fact that the “flow payoffs”  $\theta_t + e^E - \psi(e^E) - (1 - \delta)U^o$  and  $\hat{W}^E$  are uniformly bounded on  $A^E$ , then permit one to show via a standard verification argument that the constructed policy is indeed optimal and that  $W^E = \hat{W}^E$ .<sup>45</sup> ■

**Proof of Proposition 2.** Consider the linear reward scheme  $s = (s_t : \Theta^t \times \mathbb{R} \rightarrow \mathbb{R})_{t=1}^\infty$  with  $s_t(\theta^t, \pi_t) = S_t(\theta^t) + \alpha_t(\theta^t)\pi_t$ , and where  $\alpha_t(\theta^t)$  is as in (7) and  $S_t(\theta^t)$  is as in (8). Note that, because retention does not depend on cash flows, it does not affect the agent’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 agent is still employed in period  $t \geq 1$  (i.e.,  $\kappa_{t-1}(\hat{\theta}^{t-1}) = 1$ ) and any period- $t$  type  $\theta_t$ , the agent’s continuation payoff at the beginning of period  $t$  when the agent 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<sup>46</sup>

$$u_t(\theta_t; \hat{\theta}^{t-1}) \equiv \int_{\theta_t}^{\theta_t} \mathbb{E}_{\hat{\theta}_{>t}^\infty | s} \left[ \sum_{k=t}^{\tau(\hat{\theta}^{t-1}, s, \hat{\theta}_{>t}^\infty)} \delta^{k-t} J_t^k \left( s, \hat{\theta}_{>t}^k \right) \psi'(\xi_k(\hat{\theta}^{t-1}, s, \hat{\theta}_{>t}^k)) \right] ds.$$

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

Consider then Part (i). Consider an arbitrary history of past reports  $\hat{\theta}^{t-1}$ . Suppose that, in period  $t$ , the agent’s true type is  $\theta_t$  and he reports  $\hat{\theta}_t$ , 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

<sup>45</sup>This verification is standard in dynamic programming and hence omitted for brevity; the proof is however available upon request.

<sup>46</sup>Note that under the proposed scheme, the agent’s continuation payoff depends on past announcements  $\hat{\theta}^{t-1}$ , but not on past types  $\theta^{t-1}$ , effort choices  $e^{t-1}$ , or profits  $\pi^{t-1}$ .

verify that, under the proposed linear scheme, the agent's continuation payoff is then given by

$$\begin{aligned} \hat{u}_t(\theta_t, \hat{\theta}_t; \hat{\theta}^{t-1}) &= u_t(\hat{\theta}_t; \hat{\theta}^{t-1}) + \psi'(\xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t))[\theta_t - \hat{\theta}_t] \\ &\quad + \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ u_{t+1}(\tilde{\theta}_{t+1}; \hat{\theta}^{t-1}, \hat{\theta}_t) \right] - \mathbb{E}_{\tilde{\theta}_{t+1}|\hat{\theta}_t} \left[ u_{t+1}(\tilde{\theta}_{t+1}; \hat{\theta}^{t-1}, \hat{\theta}_t) \right]. \end{aligned}$$

The single-crossing conditions in the proposition then imply that, for all  $t$ , all  $\hat{\theta}^{t-1} \in \Theta^{t-1}$ , all  $\theta_t, \hat{\theta}_t \in \Theta_t$ ,

$$\left[ \frac{du_t(\theta_t; \hat{\theta}^{t-1})}{d\theta_t} - \frac{\partial \hat{u}_t(\theta_t, \hat{\theta}_t; \hat{\theta}^{t-1})}{\partial \theta_t} \right] [\theta_t - \hat{\theta}_t] \geq 0.$$

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 (see also the proof of Proposition 3 in Garrett and Pavan (2009) for similar arguments).

Now, to establish the result in the proposition, it suffices to compare the agent'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 agent'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

$$\psi'(\xi_s(\hat{\theta}^s)) = \alpha_s(\hat{\theta}^s). \tag{16}$$

Restricting attention to continuation strategies in which, at any period  $s \geq t$ , the agent 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 agent'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 (16), there exists another continuation strategy whose effort choices comply with (16) for all  $s \geq t$ , all  $\hat{\theta}^s$ , which gives the agent a (weakly) higher expected continuation payoff.



Next, it is easy to see that, under any continuation strategy that satisfies the aforementioned effort property, the agent'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  $\theta_t \in \Theta_t$  all  $\hat{\theta}^{t-1}, \check{\theta}^{t-1} \in \Theta^{t-1}$ ,  $\delta^t \left| \hat{u}_t(\theta_t; \hat{\theta}^{t-1}) - \check{u}_t(\theta_t; \check{\theta}^{t-1}) \right| < \epsilon$ , where  $\hat{u}_t$  and  $\check{u}_t$  are continuation payoffs under arbitrary continuation strategies satisfying the above effort restriction, given arbitrary histories of reports  $\hat{\theta}^{t-1}$  and  $\check{\theta}^{t-1}$ .

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 agent's continuation payoff under a truthful and obedient strategy from that period onwards is weakly higher than the expected payoff under *any* alternative continuation strategy. We thus conclude that, whenever the pair  $(\xi, \kappa)$  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 (11). 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 in which the policies  $(\xi^*, \kappa^*)$  – now defined over  $\Theta$  – 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 (11) is independent of the retention policy  $\kappa$  and is such that  $\xi_t^*(\theta^t)$  is given by (12), all  $t$  all  $\theta^t \in R^t$ . Next, consider the retention policy. We first prove existence of a unique fixed point  $W^* \in \mathcal{B}$  to the mapping  $T$  defined in the proposition. To this end, endow  $\mathcal{B}$  with the uniform metric. That  $\mathcal{B}$  is closed under  $T$

is ensured by the restrictions on  $\psi$  and by the definition of  $\xi^*$ , which together imply that each function  $VS_t : R^t \rightarrow \mathbb{R}$  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) implies the result. Standard arguments then permit one to verify that  $W^*(\theta^t, t)$  is indeed the value associated with the problem that consists of choosing a retention policy from period  $t$  onwards so as to maximize (11), given the history of productivities  $\theta^t \in R^t$  and given the profit-maximizing effort policy  $\xi^*$ . Having established this result, it is then immediate that any policy  $\kappa^*$  that, given the effort policy  $\xi^*$ , maximizes (11) 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  $\underline{\theta}_1$  obtains an expected payoff under  $\Omega^*$  equal to his outside option, i.e.,  $V^{\Omega^*}(\underline{\theta}_1) = U^o$ ; and (iii) the agent never finds it optimal to leave the firm when offered the possibility of staying. That the mechanism  $\Omega^*$  is optimal then follows directly from the fact that the firm's ex-ante expected profits from hiring a new agent are given by (11) in any mechanism that is incentive compatible and individually rational for the agent.

*Step 3.* Now assume that each function  $h_t(\cdot) \equiv -\eta(\cdot)J_1^t(\cdot)$  is nondecreasing. Because the function  $g(e, h, \theta) \equiv e + \theta - \psi(e) + h\psi'(e)$  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  $-\eta(\cdot)J_1^t(\cdot)$  is nondecreasing, so is the function  $VS_t(\cdot)$  — this is an immediate implication of the envelope theorem. This property, together with the fact that the process describing the evolution

of the agents' productivities satisfies the assumption of "first-order stochastic dominance in types" implies that  $\mathcal{N}$  is closed under the operator  $T$ . It follows that  $\lim_{n \rightarrow \infty} T^n W$  is in  $\mathcal{N}$ . The fact that  $T : \mathcal{B} \rightarrow \mathcal{B}$  admits a unique fixed point then implies that  $\lim_{n \rightarrow \infty} T^n W = W^*$ .

The last result, together with "first-order stochastic dominance in types" implies that, for each  $t$ , each  $\theta^{t-1} \in R^{t-1}$ ,  $\mathbb{E}_{\tilde{\theta}^{t+1} | (\theta^{t-1}, \cdot)} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right]$  is nondecreasing in  $\theta_t$ . Given the monotonicity of each function  $\mathbb{E}_{\tilde{\theta}^{t+1} | (\theta^{t-1}, \cdot)} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right]$ , it is then immediate that the retention policy  $\kappa^*$  that maximizes (11) 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  $\theta^{t-1} \in R^{t-1}$ ,

$$\theta_t^*(\theta^{t-1}) = \begin{cases} -K & \text{if } \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right] > \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right] \text{ for all } \theta_t \text{ s.t. } (\theta^{t-1}, \theta_t) \in R^t \\ K & \text{if } \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right] < \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right] \text{ for all } \theta_t \text{ s.t. } (\theta^{t-1}, \theta_t) \in R^t \\ \min \left\{ \theta_t \in \Theta_t : (\theta^{t-1}, \theta_t) \in R^t \text{ and } \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^* \left( \tilde{\theta}^{t+1}, t+1 \right) \right] \geq \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right] \right\} & \text{otherwise} \end{cases}$$

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  $\varphi_s : \Theta^t \rightarrow \Theta^t$  be the function defined, for all  $\theta^t \in \Theta^t$ , by

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

For all  $\theta_1 \in \Theta_1$ , let  $\lambda_1(\theta_1) = \theta_1$ , while for any  $t \geq 2$ , let  $\lambda_t : \Theta^t \rightarrow R^t$  be the function defined, for all  $\theta^t \in \Theta^t$ , by  $\lambda_t(\theta^t) \equiv \varphi_t \circ \varphi_{t-1} \circ \dots \circ \varphi_2(\theta^t)$ . Note that the function  $\lambda_t$  maps each vector of reports  $\theta^t \in \Theta^t \setminus R^t$  which reveals that the agent 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  $\bar{\xi}^*$  and  $\bar{\kappa}^*$  be the policies defined by  $\bar{\xi}_t^*(\theta^t) = \xi_t^*(\lambda_t(\theta^t))$  and  $\bar{\kappa}_t^*(\theta^t) = \kappa_t^*(\lambda_t(\theta^t))$ , 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  $\bar{\xi}_t^*(\cdot)$  and  $\bar{\kappa}_t^*(\cdot)$ . This in turn guarantees that the extended policies  $\bar{\xi}^* = (\bar{\xi}_t^*)_{t=1}^\infty$  and  $\bar{\kappa}^* = (\bar{\kappa}_t^*)_{t=1}^\infty$  satisfy all the single-crossing conditions of Proposition 2. ■

**Proof of Example 1.** It is easy to see that (i) the supports  $\Theta_t$  of the marginal distributions are

intervals; (ii) that given any  $t \geq 1$ , any  $\theta_{t-1} \in \Theta_{t-1}$ , the conditional distribution  $F_t(\cdot|\theta_{t-1})$  is absolutely continuous over  $\mathbb{R}$  with support a connected subset of  $\Theta_t$ ; (iii) for any  $t \geq 2$ , any  $\theta_t \in \mathbb{R}$ , almost any  $\theta_{t-1} \in \Theta_{t-1}$ ,  $\partial F_t(\theta_t|\theta_{t-1})/\partial\theta_{t-1}$  exists; (iv) all kernels  $F_t(\cdot|\theta^{t-1})$  satisfy the condition of first-order stochastic dominance in types.

To see that the sets  $\Theta_t$  are bounded uniformly over  $t$ , let  $\bar{\varepsilon} \in \mathbb{R}_{++}$  be the uniform upper bound on  $\varepsilon_t$ , i.e.,  $|\varepsilon_t| \leq \bar{\varepsilon}$  all  $t$  all  $\varepsilon_t \in [\underline{\varepsilon}_t, \bar{\varepsilon}_t]$ . Then, for each  $t$ , the upper and lower bounds on the marginal distribution  $\Theta_t$  satisfy

$$|\underline{\theta}_t|, |\bar{\theta}_t| \leq \times_{\tau=2}^t \gamma_\tau \max\{|\underline{\theta}_1|, |\bar{\theta}_1|\} + \sum_{k=1}^t (\times_{\tau=k+1}^t \gamma_\tau) \bar{\varepsilon}.$$

The condition that  $\sum_{k=1}^t (\times_{\tau=k+1}^t \gamma_\tau)$  is bounded uniformly across  $t$  then gives the result.

Next note that, for any  $t$  and  $\tau$ ,  $\tau > t$ , any  $\theta^\tau \in R^\tau$ ,  $J_t^\tau(\theta_{t-1}^\tau) = \times_{k=t+1}^\tau \gamma_k$  which is also bounded uniformly, since  $\sum_{k=1}^t (\times_{\tau=k+1}^t \gamma_\tau)$  is bounded uniformly.

Finally, it is easy to see that, for any  $t \geq 2$ , any  $\theta^t \in R^t$ ,  $\eta(\theta_1) J_1^t(\theta^t) = \eta(\theta_1) \times_{k=2}^t \gamma_k$  is nonincreasing in  $\theta_1$  and constant in  $\theta_{>1}^t$ , so that the monotonicity condition of Proposition 3 in the main text is satisfied. ■

**Proof of Example 2.** Note that, for any  $t \geq 2$ , any  $\theta_t \in \mathbb{R}$  any  $\theta_{t-1} \in \Theta_{t-1}$ ,

$$\begin{aligned} F_t(\theta_t|\theta_{t-1}) &= \Pr(\bar{\theta} - \varphi(\theta_{t-1}) \tilde{\varepsilon}_t \leq \theta_t) \\ &= \Pr\left(\tilde{\varepsilon}_t \geq \frac{\bar{\theta} - \theta_t}{\varphi(\theta_{t-1})}\right) \\ &= 1 - G_t\left(\frac{\bar{\theta} - \theta_t}{v(\theta_{t-1})}\right). \end{aligned}$$

Given the assumptions on  $F_1$ ,  $\varphi(\cdot)$ , and  $G_t$ , it is then easy to see that (i) the supports  $\Theta_t$  of the marginal distributions are intervals with  $\Theta_t \subset \Theta'$  all  $t$ ; (ii) given any  $t \geq 1$ , any  $\theta_{t-1} \in \Theta_{t-1}$ , the conditional distribution  $F_t(\cdot|\theta_{t-1})$  is absolutely continuous over  $\mathbb{R}$  with support a connected subset of  $\Theta_t$ ; (iii) for any  $t \geq 2$ , any  $\theta_t \in \mathbb{R}$ , almost any  $\theta_{t-1} \in \Theta_{t-1}$ ,  $\partial F_t(\theta_t|\theta_{t-1})/\partial\theta_{t-1}$  exists; (iv) all kernels  $F_t(\cdot|\theta^{t-1})$  satisfy the condition of first-order stochastic dominance in types. Furthermore, for all  $k \geq 2$ ,

$\theta_{k-1} \in \Theta_{k-1}$ , and  $\theta_k \in \Theta_k$

$$\begin{aligned}\frac{\partial F_k(\theta_k|\theta_{k-1})}{\partial \theta_{k-1}} &= \frac{\varphi'(\theta_{k-1})(\bar{\theta} - \theta_k)}{\varphi(\theta_{k-1})^2} g_k\left(\frac{\bar{\theta} - \theta_k}{\varphi(\theta_{k-1})}\right) \\ f_k(\theta_k|\theta_{k-1}) &= \frac{1}{\varphi(\theta_{k-1})} g_k\left(\frac{\bar{\theta} - \theta_k}{\varphi(\theta_{k-1})}\right) \\ I_{k-1}^k(\theta_{k-1}, \theta_k) &= \frac{-\varphi'(\theta_{k-1})(\bar{\theta} - \theta_k)}{\varphi(\theta_{k-1})}\end{aligned}$$

or, equivalently,

$$I_{k-1}^k(\theta_{k-1}, \theta_k) = -\varphi'(\theta_{k-1}) \varepsilon_k$$

for some  $\varepsilon_k \in [0, 1]$ . Since  $\varphi'(\theta) \in [-1, 0]$  for all  $\theta \in \Theta'$ , the impulse responses  $J_t^\tau(\theta_{>t-1}^\tau) \equiv \Pi_{k=t+1}^\tau I_{k-1}^k(\theta_k, \theta_{k-1})$ , are bounded uniformly by 1. The process thus satisfies all the properties of the model set up.

Lastly, note that since  $\varphi'(\cdot)/\varphi(\cdot)$  is nondecreasing, and since, for all  $k \geq 2$ ,  $\theta_{k-1} \in \Theta_{k-1}$ , and  $\theta_k \in \Theta_k$ ,  $\bar{\theta} - \theta_k \geq 0$  and  $\varphi'(\theta_{k-1})/\varphi(\theta_{k-1}) \leq 0$ ,  $I_{k-1}^k(\cdot)$  is nonincreasing. Together with the assumption that  $\eta(\cdot)$  is nonincreasing, this implies that, for any  $t$ , any  $\theta^t \in R^t$ ,  $\eta(\cdot)J_1^t(\cdot)$  is nonincreasing, which completes the proof. ■

**Proof of Proposition 4.** Let  $\hat{\mathcal{N}}$  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  $R^t$ ; and (b) for any  $s > t$ , any  $\theta^t \in R^t$  and any  $\theta_{>t}^s$  such that  $\theta^s = (\theta^t, \theta_{>t}^s) \in R^s$ , 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  $T: \mathcal{B} \rightarrow \mathcal{B}$  thus satisfies properties (a) and (b) above. First-order stochastic dominance in types then implies that

$$\mathbb{E}_{\tilde{\theta}^{s+1}|\theta^s} \left[ W^*(\tilde{\theta}^{s+1}, s+1) \right] \geq \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t} \left[ W^*(\tilde{\theta}^{t+1}, t+1) \right].$$

■

**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 (1).* 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 outside option.

**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  $T_{E,c} : \mathcal{B}^E \rightarrow \mathcal{B}^E$  defined, for all  $W \in \mathcal{B}^E$ , all  $(\theta_t, t) \in A^E$ , by*

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

*Fix  $c', c'' \in \mathbb{R}$  with  $c'' > c'$ . There exists  $\iota > 0$  such that, for all  $t$ , all  $\theta_t \in \Theta_t$ ,  $\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W^{E,c''}(\tilde{\theta}_{t+1}, t+1) \right] \geq c''$  implies  $\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W^{E,c'}(\tilde{\theta}_{t+1}, t+1) \right] > c' + \iota$ .*

**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  $A^E$  to  $\mathbb{R}$  such that, for all  $(\theta_t, t) \in A^E$ ,  $W(\theta_t, t) \geq W^{E,c''}(\theta_t, t) - \delta(c'' - c')$ . 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  $(\theta_t, t)$ ,

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

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  $W^{E,c'} \in \mathcal{B}^E$  to the operator  $T_{E,c'}$  must be an element of  $\mathcal{C}(c'', c')$ . That is, for all  $(\theta_t, t)$ ,  $W^{E,c'}(\theta_t, t) - W^{E,c''}(\theta_t, t) \geq -\delta(c'' - c')$ .

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

$$\begin{aligned} \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W^{E,c'}(\tilde{\theta}_{t+1}, t+1) \right] &\geq \mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t} \left[ W^{E,c''}(\tilde{\theta}_{t+1}, t+1) \right] - \delta(c'' - c') \\ &\geq c'' - \delta(c'' - c') \\ &> c' + \iota \end{aligned}$$

for some  $\iota > 0$ . ■

The next lemma establishes a strict ranking between the endogenous outside options in the stopping problems corresponding to the efficient and profit-maximizing dynamic programs.

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

**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  $W (\theta^t, t) \leq W^E (\theta_t, t)$  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  $(\theta^t, t) \in A$ ,

$$\begin{aligned} TW (\theta^t, t) &= \xi_t^* (\theta^t) + \theta_t - \psi (\xi_t^* (\theta^t)) - \eta (\theta_1) J_1^t (\theta^t) \psi' (\xi_t^* (\theta^t)) - (1 - \delta) U^o \\ &\quad + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W \left( \tilde{\theta}^t, t + 1 \right) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W \left( \tilde{\theta}_1, 1 \right) \right] \right\} \\ &\leq e^E + \theta_t - \psi (e^E) - (1 - \delta) U^o \\ &\quad + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}_{t+1} | \theta_t} \left[ W^E \left( \tilde{\theta}_{t+1}, t + 1 \right) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W^E \left( \tilde{\theta}_1, 1 \right) \right] \right\} \\ &= T_E W^E (\theta_t, t) \\ &= W^E (\theta_t, t). \end{aligned}$$

Since  $\mathcal{D} (W^E)$ , together with the uniform metric, is a complete metric space, and since  $T$  is a contraction, given any  $W \in \mathcal{D} (W^E)$ ,  $\lim_{n \rightarrow \infty} T^n W$  exists and belongs to  $\mathcal{D} (W^E)$ . Since  $W^*$  is the unique fixed point to the mapping  $T : \mathcal{B} \rightarrow \mathcal{B}$ , it must be that  $W^* = \lim_{n \rightarrow \infty} T^n W$ . 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  $\theta_1 \in \Theta_1 \setminus \{\bar{\theta}_1\}$ ,

$$\begin{aligned} W^* (\theta_1, 1) &= TW^* (\theta_1, 1) \\ &= \xi_1^* (\theta_1) + \theta_1 - \psi (\xi_1^* (\theta_1)) - \eta (\theta_1) \psi' (\xi_1^* (\theta_1)) - (1 - \delta) U^o \\ &\quad + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^2 | \theta_1} \left[ W^* \left( \tilde{\theta}^2, 2 \right) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right] \right\} \\ &< \theta_1 + e^E - \psi (e^E) - (1 - \delta) U^o \\ &\quad + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}_2 | \theta_1} \left[ W^E \left( \tilde{\theta}_2, 2 \right) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W^E \left( \tilde{\theta}_1, 1 \right) \right] \right\} \\ &= W^E (\theta_1, 1), \end{aligned}$$

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

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

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

$$\mathbb{E}_{\tilde{\theta}_{t+1}|\theta^t} \left[ W^E \left( \tilde{\theta}_{t+1}, t+1 \right) \right] \geq \mathbb{E}_{\tilde{\theta}_1} \left[ W^E \left( \tilde{\theta}_1, 1 \right) \right]$$

*implies*

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

**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  $\theta_t + e^E - \psi(e^E) - (1-\delta)U^o$  and exogenous outside option  $c'$ . Now let  $c' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right]$ . 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

$$VS_t(\theta^t) \equiv \xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1)J_1^t(\theta^t) \psi'(\xi_t^*(\theta^t)) - (1-\delta)U^o$$

and outside option  $c' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right]$ . By the property of “vanishing impulse responses”, for any  $\omega > 0$ , there exists  $\bar{t}$  such that, for any  $t > \bar{t}$ , any  $\theta^t \in R^t$ ,

$$VS_t(\theta^t) > \theta_t + e^E - \psi(e^E) - (1-\delta)U^o - \omega.$$

That is, for  $t > \bar{t}$ , 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 outside option  $c' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right]$ . In terms of value functions, this implies that, for all  $t > \bar{t}$ , all  $\theta^t \in R^t$ ,

$$W^*(\theta^t, t) \geq W^{E,c'}(\theta^t, t) - \frac{\omega}{1-\delta}. \tag{17}$$

To see this, consider the set  $\mathcal{W} \subset \mathcal{B}$  of all bounded functions  $W$  from  $A$  to  $\mathbb{R}$  such that, for all  $t > \bar{t}$ , 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  $T_{c'} : \mathcal{B} \rightarrow \mathcal{B}$  defined, for all  $(\theta^t, t)$ , by

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



The set  $\mathcal{W}$  is closed under  $T_{c'}$ . Indeed, if  $W \in \mathcal{W}$ , then, for any  $t > \bar{t}$ , any  $\theta^t \in R^t$ ,

$$\begin{aligned} T_{c'} W(\theta^t, t) - W^{E, c'}(\theta_t, t) &= V S_t(\theta^t) + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W(\tilde{\theta}^{t+1}, t+1) \right], c' \right\} \\ &\quad - \left( \theta_t + e^E - \psi(e^E) - (1-\delta)U^o \right. \\ &\quad \left. + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^{E, c'}(\tilde{\theta}_{t+1}, t+1) \right], c' \right\} \right) \\ &\geq -\omega - \frac{\delta\omega}{1-\delta} = -\frac{\omega}{1-\delta}. \end{aligned}$$

Since  $\mathcal{W}$ , together with the uniform metric, is a complete metric space, and since  $T_{c'}$  is a contraction, given any  $W \in \mathcal{W}$ ,  $\lim_{n \rightarrow \infty} T_{c'}^n W$  exists and belongs to  $\mathcal{W}$ . Furthermore, because  $c' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^*(\tilde{\theta}_1, 1) \right]$ , it must be that  $W^* = \lim_{n \rightarrow \infty} T_{c'}^n W$ . Hence,  $W^* \in \mathcal{W}$ , which proves (17).

Now, let  $c'' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^E(\tilde{\theta}_1, 1) \right]$ . By Lemma A2,  $c'' > c'$ . Now observe that  $W^E = W^{E, c''}$ . It follows that, for all  $t > \bar{t}$  and all  $\theta^t \in R^t$ , if  $\mathbb{E}_{\tilde{\theta}_{t+1} | \theta^t} \left[ W^E(\tilde{\theta}_{t+1}, t+1) \right] \geq \mathbb{E}_{\tilde{\theta}_1} \left[ W^E(\tilde{\theta}_1, 1) \right]$ , then

$$\begin{aligned} \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W^*(\tilde{\theta}^{t+1}, t+1) \right] &\geq \mathbb{E}_{\tilde{\theta}_{t+1} | \theta^t} \left[ W^{E, c'}(\tilde{\theta}_{t+1}, t+1) \right] - \frac{\omega}{1-\delta} \\ &> \mathbb{E}_{\tilde{\theta}_1} \left[ W^*(\tilde{\theta}_1, 1) \right] + \iota - \frac{\omega}{1-\delta}. \end{aligned}$$

The first inequality follows by (17), and the second by Lemma A1 using  $c' = \mathbb{E}_{\tilde{\theta}_1} \left[ W^*(\tilde{\theta}_1, 1) \right]$  and choosing  $\iota$  as in that lemma. The result then follows by choosing  $\omega$  sufficiently small that  $\iota - \frac{\omega}{1-\delta} > 0$ .

■

*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  $\eta(\theta_1) J_1^t(\theta_1, \cdot)$  is Lipschitz continuous over  $\Theta_{>1}^t(\theta_1) \equiv \{\theta_{>1}^t \in \Theta_{>1}^t : (\theta_1, \theta_{>1}^t) \in R^t\}$  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  $f_t(\theta_t | \cdot)$  is Lipschitz continuous over  $\Theta_{t-1}$  with Lipschitz constant  $\rho$ . Then, for each  $t \geq 2$  and each  $\theta^{t-1} \in R^{t-1}$ ,  $\mathbb{E}_{\tilde{\theta}^{t+1} | (\theta^{t-1}, \cdot)} \left[ W^*(\theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1) \right]$  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  $t \geq 2$ ,  $\mathbb{E}_{\tilde{\theta}_{t+1}} \left[ W^E(\tilde{\theta}_{t+1}, t+1) \right]$  is Lipschitz continuous over  $\Theta_t$ .<sup>47</sup>*

<sup>47</sup>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  $\theta^{t-1} \in R^{t-1}$ ,  $\mathbb{E}_{\tilde{\theta}^{t+1} | (\theta^{t-1}, \cdot)} \left[ W^* \left( \theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1 \right) \right]$  is Lipschitz continuous in  $\Theta_t(\theta^{t-1})$ . The proof that  $\mathbb{E}_{\tilde{\theta}_{t+1}} \left[ W^E \left( \tilde{\theta}_{t+1}, t+1 \right) \right]$  is Lipschitz continuous on  $\Theta_t$  is similar and omitted.

Let

$$M \equiv \frac{e^E - \psi(e^E) + K}{1 - \delta} + U^o$$

and define

$$m \equiv \frac{1 + \beta L + 2\delta\rho MK}{1 - \delta},$$

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  $\Theta_{>1}^t(\theta_1)$  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  $\Theta^t(\theta_1)$  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  $(\theta^t, t) \in A$ ,

$$\begin{aligned} TW(\theta^t, t) &= VS_t(\theta^t) + \delta \max \left\{ \mathbb{E}_{\tilde{\theta}^{t+1} | \theta^t} \left[ W \left( \tilde{\theta}^{t+1}, t+1 \right) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W \left( \tilde{\theta}_1, 1 \right) \right] \right\} \\ &\leq e^E + K - \psi(e^E) - (1 - \delta)U^o + \delta M \\ &= M - 2(1 - \delta)U^o \leq M. \end{aligned}$$

Next note that, for any  $(\theta^t, t) \in A$ ,  $TW(\theta^t, t) \geq -K - (1 - \delta)U^o - \delta M > -M$ . 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  $(\theta^{\tau-1}, \theta'_\tau, \theta_{>\tau}^t), (\theta^{\tau-1}, \theta''_\tau, \theta_{>\tau}^t) \in R^t$ .

Suppose, without loss of generality, that  $\theta'_\tau > \theta''_\tau$ . Then,

$$\begin{aligned}
& TW(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, t) - TW(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, t) \\
&= [\xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1)J_1^t(\theta^t) \psi'(\xi_t^*(\theta^t))]_{\theta^t=(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})} \\
&\quad - [\xi_t^*(\theta^t) + \theta_t - \psi(\xi_t^*(\theta^t)) - \eta(\theta_1)J_1^t(\theta^t) \psi'(\xi_t^*(\theta^t))]_{\theta^t=(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \\
&\quad + \delta \left( \begin{aligned} & \max \left\{ \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W(\tilde{\theta}_1, 1) \right] \right\} \\ & - \max \left\{ \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right], \mathbb{E}_{\tilde{\theta}_1} \left[ W(\tilde{\theta}_1, 1) \right] \right\} \end{aligned} \right)
\end{aligned} \tag{18}$$

The first two terms on the right-hand side of (18) are no greater than  $(1 + \beta L)(\theta'_\tau - \theta''_\tau)$ . This can be derived as follows. For any  $2 \leq \tau \leq t$ , any  $(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) \in R^t$ , define  $\Theta_\tau(\theta^{\tau-1}, \theta^t_{>\tau}) \equiv \{\theta_\tau \in \Theta_\tau : (\theta^{\tau-1}, \theta_\tau, \theta^t_{>\tau}) \in R^t\}$ . Define, for any  $(\theta^{\tau-1}, \theta_\tau, \theta^t_{>\tau}) \in R^t$  and any  $e \in E$ , the flow virtual surplus function

$$g_t(\theta^t, e) = e + \theta_t - \psi(e) - \eta(\theta_1)J_1^t(\theta^t) \psi'(e).$$

For any  $\theta^t = (\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) \in R^t$ ,  $g_t$  is Lipschitz continuous in  $\theta_\tau$  and  $\frac{\partial}{\partial \theta_\tau} g_t(\theta^{\tau-1}, \theta_\tau, \theta^t_{>\tau}, e) \leq 1 + \beta L$  for all  $e \in E$  and almost all  $\theta_\tau \in \Theta_\tau(\theta^{\tau-1}, \theta^t_{>\tau})$ . 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 (18) is no greater than  $\delta(2\rho MK + m)(\theta'_\tau - \theta''_\tau)$ . This follows because

$$\begin{aligned}
& \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right] \\
& - \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right] \\
&= \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right] \\
& - \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \left[ W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \right] \\
& + \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \left( \begin{aligned} & W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \\ & - W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \end{aligned} \right) \\
&= \int_{\Theta_{t+1}} W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \theta_{t+1}, t+1) \left( \begin{aligned} & f_{t+1}(\theta_{t+1} | \theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) \\ & - f_{t+1}(\theta_{t+1} | \theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}) \end{aligned} \right) d\theta_{t+1} \\
& + \mathbb{E}_{\tilde{\theta}_{t+1} | (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \left( \begin{aligned} & W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \\ & - W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde{\theta}_{t+1}, t+1) \end{aligned} \right) \\
&\leq (2\rho MK + m)(\theta'_\tau - \theta''_\tau),
\end{aligned} \tag{19}$$

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

$$\begin{aligned} TW(\theta^{\tau-1}, \theta'_\tau, \theta_{>\tau}^t, t) - TW(\theta^{\tau-1}, \theta''_\tau, \theta_{>\tau}^t, t) &\leq (1 + \beta L + 2\delta\rho MK + \delta m)(\theta'_\tau - \theta''_\tau) \\ &= m(\theta'_\tau - \theta''_\tau). \end{aligned}$$

Since  $(\theta^{\tau-1}, \theta'_\tau, \theta_{>\tau}^t)$  and  $(\theta^{\tau-1}, \theta''_\tau, \theta_{>\tau}^t)$  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  $\Theta^t(\theta_1)$  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 already argued 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 (19), we then have that  $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^{t-1}, \cdot} \left[ W^* \left( \theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1 \right) \right]$  is Lipschitz continuous over  $\Theta_t(\theta^{t-1})$  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  $\bar{t}$  be as defined in Lemma A3. Take an arbitrary  $t > \bar{t}$  and  $\theta^{t-1} \in R^{t-1}$  such that  $\theta_t^E \in \text{int}\{Supp[F_t(\cdot|\theta_{t-1})]\}$ . The continuity of  $\mathbb{E}_{\tilde{\theta}_{t+1}|\cdot} \left[ W^E \left( \tilde{\theta}_{t+1}, t+1 \right) \right]$  in  $\theta_t$  established in the previous lemma, implies

$$\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t^E} \left[ W^E \left( \tilde{\theta}_{t+1}, t+1 \right) \right] = \mathbb{E}_{\tilde{\theta}_1} \left[ W^E \left( \tilde{\theta}_1, 1 \right) \right].$$

Since  $t > \bar{t}$ , by Lemma A3, it follows that

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

By Lemma A4,  $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^{t-1}, \cdot} \left[ W^* \left( \theta^{t-1}, \cdot, \tilde{\theta}_{t+1}, t+1 \right) \right]$  is continuous in  $\theta_t$ . Since  $\theta_t^E \in \text{int}\{Supp[F_t(\cdot|\theta_{t-1})]\}$ , there exists  $\epsilon > 0$  such that, for all  $\theta_t \in (\theta_t^E - \epsilon, \theta_t^E)$ ,

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

and  $(\theta^{t-1}, \theta_t) \in R^t$ . It follows that  $\theta_t^* (\theta^{t-1}) < \theta_t^E$ . ■ □

**Proof of Proposition 6.** Because the process is autonomous  $\theta_t^E = \theta^E$ , all  $t$ . Firstly, suppose that  $\theta^E < \bar{\theta}_1$ . Consider the case that, for all  $\theta_1 > \theta^E$ ,

$$\mathbb{E}_{\bar{\theta}_2|\theta_1} \left[ W^* \left( \tilde{\theta}^2, 2 \right) \right] > \mathbb{E} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right].$$

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  $\theta_1, \theta_t > \theta^E$ ,

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

This means that, for any  $t$  any  $\theta^t \in R^t$  such that  $\kappa_{t-1}^*(\theta^{t-1}) = 1$  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 agent 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

$$\mathbb{E}_{\bar{\theta}_2|\theta_1} \left[ W^* \left( \tilde{\theta}^2, 2 \right) \right] < \mathbb{E} \left[ W^* \left( \tilde{\theta}_1, 1 \right) \right],$$

which implies that  $\theta_1^* > \theta^E$ . By assumption, the agent is retained with positive probability after the first period, i.e.  $\theta_1^* \in (\theta_1^E, \bar{\theta})$ . The result in Part (a) then holds by letting  $\underline{t} = 2$ , whereas the result in Part (b) follows from Part (i) of Proposition 5. Indeed, from that proposition, there exists a  $\bar{t} > \underline{t}$  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 consider the case that  $\theta^E = \bar{\theta}_1$ . Case (i) then always trivially applies. ■

**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 agent in period 2 is

$$\mathbb{E} \left[ \xi_1^* \left( \tilde{\theta}_1 \right) + \tilde{\theta}_1 - \psi(\xi_1^*(\tilde{\theta}_1)) - \eta(\tilde{\theta}_1)\psi'(\xi_1^*(\tilde{\theta}_1)) \right] = \frac{1}{6}.$$

The agent is thus retained if and only if

$$\mathbb{E}_{\tilde{\theta}_2|\theta_1} \left[ \xi_2^*(\theta_1) + \tilde{\theta}_2 - \psi(\xi_2^*(\theta_1)) - \eta(\theta_1)\gamma\psi'(\xi_2^*(\theta_1)) \right] \geq \frac{1}{6},$$

where  $\xi_2^*(\theta_1) = 1 - \frac{\gamma}{2} + \gamma\theta_1$  and  $\mathbb{E}_{\tilde{\theta}_2|\theta_1} [\tilde{\theta}_2] = \gamma\theta_1$ . 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  $\gamma > 0.845$ . ■