Managerial Turnover in a Changing World

Daniel F. Garrett

Toulouse School of Economics

Alessandro Pavan

Northwestern University

We develop a dynamic theory of managerial turnover in a world in which the quality of the match between a firm and its managers changes stochastically over time. Shocks to managerial productivity are anticipated at the time of contracting but privately observed by the managers. Our key positive result shows that the firm's optimal retention decisions become more permissive with time. Our key normative result shows that, compared to what is efficient, the firm's contract induces either excessive retention at all tenure levels or excessive firing at the early stages of the relationship, followed by excessive retention after sufficiently long tenure.

I. Introduction

The job security and pay of a firm's top manager typically rest 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

For useful comments and suggestions, we thank the editor, Phillip Reny, two anonymous referees, Igal Hendel, Jin Li, Alessandro Lizzeri, Robert Miller, Dale Mortensen, Marco Ottaviani, Bill Rogerson, and seminar participants at Berkeley, Harvard, Massachusetts Institute of Technology, Northwestern, Princeton, Stanford, Yale, the European University Institute, the 2010 Gerzensee European Symposium in Economic Theory, the 2010 Econometric Society winter meetings, and the 2010 Econometric Society World Congress.

[*Journal of Political Economy*, 2012, vol. 120, no. 5]

so in the future.[1] Shocks to managerial productivity may originate from the opening of new markets, the arrival of new technologies, industry consolidation, or the introduction of new legislation.

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

In this paper, we develop a dynamic theory of managerial contracting that, in addition to the familiar theme of incentivizing effort, accounts explicitly for the following possibilities: (i) managerial ability to generate profits is bound to change (stochastically) over time; (ii) shocks to managerial productivity are anticipated at the time of contracting but privately observed by the managers; (iii) at each point in time, the board can respond to poor future prospects by replacing an incumbent manager with a new hire; and (iv) the firm's performance under each new hire is going to be affected by the same information frictions as in the relationship with the incumbent.

Not only is accounting for these possibilities realistic, but it sheds new light on the joint dynamics (and inefficiency) of effort, retention, and compensation decisions.

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

Upon separating from the incumbent, the firm goes back to the labor market and is randomly matched with a new manager of unknown productivity. Each manager's initial productivity (i.e., his productivity at the time of contracting) is his own private information. 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

[1] For example, Fine (1998) argues that technology is increasing the speed at which business environments evolve across a plethora of industries.

environment with the tasks that the manager is asked to perform affects the evolution of his productivity. The environment is perfectly stationary in the sense that the firm faces the same problem with each manager it hires. As a result, the board offers the same menu of contracts to each manager.[2]

A contract is described by (i) an effort policy specifying in each period the effort recommended to the manager, (ii) a retention policy specifying in each period whether the manager will be retained in the next period or permanently fired, and (iii) a compensation policy specifying in each period the manager's compensation. The first two policy functions can depend on past and current (self-)reported managerial productivity and past cash flows, while the current period's compensation policy can in addition depend on the current period's cash flow.[3]

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

[2] While our analysis focuses on a representative firm, both our positive and normative results apply also to certain competitive labor markets where, after dismissal, managers go back to the market and are randomly matched with other identical firms. What makes a policy of "selling the firm to the managers" suboptimal is the fact that the managers have private information about their abilities to generate profits for the firm. This private information, since it originates in idiosyncratic characteristics as well as past working experiences, is present from the very first moment a manager is matched with the firm and has persistent (although typically diminishing) effects over time. Because of such private information, if the firm were sold to the managers, then any type above the lowest would get the full surplus of his higher productivity. To extract some of this surplus, the board of directors instead retains control of the firm and introduces distortions in the contracts that govern managers' effort and separation decisions.

[3] In general, a turnover policy based solely on observed cash flows cannot induce the optimal sequence of separation decisions. It may be essential that managers keep communicating with the board, e.g., by explaining the determinants of past performances and/or by describing the firm's prospects under their control. A key role of the optimal contract in our theory is precisely to induce a prompt exchange of information between the managers and the board, in addition to the more familiar role of incentivizing effort through performance-based compensation.

through a fixed-point dynamic programming problem that internalizes all relevant trade-offs and whose solution endogenizes the firm's separation payoff.[4]

Key positive results.—Our key positive prediction is that the firm's optimal retention decisions become more permissive with time: the productivity level that the firm requires for each manager to be retained declines with the number of periods that the manager has been working for the firm. 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,[5] and (ii) variations in managerial productivity are anticipated but privately observed.

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

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

The firm's profit-maximizing retention policy is then obtained by trading off two considerations. The first is the desire to respond promptly and efficiently to variations in the environment that affect the firm's prospects under the incumbent's control, of course taking into account the

[4] Note that endogenizing the payoff the firm expects after separating from each incumbent manager is essential to the normative results in the paper.

[5] Below, we provide a formal statement of this assumption in terms of a statistical property of the process governing the evolution of managerial productivity.

dynamics of future effort and retention decisions: this concern calls for retaining managers whose productivity is expected to remain or turn high irrespective of whether or not their initial productivity was low. The second consideration is the value of offering a contract that reduces the compensation that the firm must pay to the managers who are most productive at the hiring stage: this second concern calls for committing to a retention policy that is most severe to those managers whose initial productivity is low. However, because the value of such commitments declines with the length of the employment relationship, the profit-maximizing retention policy becomes gradually more lenient over time.

Our theory thus offers a possible explanation for what in the eyes of an external observer may look like "entrenchment." That managers with a longer tenure are retained under the same conditions that would have called for separation at a shorter tenure is, in our theory, the result of a fully optimal contract as opposed to the result of a lack of commitment or of good governance. In this respect, our explanation is fundamentally different from the alternative view that managers with longer tenure are "entrenched" because they are able to exert more influence over the board, either because of manager-specific investments, as in Shleifer and Vishny (1989), or because of the appointment of less independent directors, as in Hermalin and Weisbach (1998); see also Weisbach (1988), Denis, Denis, and Sarin (1997), Hadlock and Lumer (1997), Rose and Shepard (1997), Almazan and Suarez (2003), Bebchuk and Fried (2004), and Fisman, Kuhrana, and Rhodes-Kropf (2005).

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

The result that retention decisions become less efficient over time may appear in contrast to findings in the dynamic mechanism design literature that "distortions" in optimal contracts typically decrease over time and vanish in the long run. (This property has been documented by various authors, going back at least to Besanko's [1985] seminal work; see Battaglini [2005] for a recent contribution and Pavan, Segal, and Toikka [2012] for a unifying explanation based on the statistical property of declining impulse responses.)

The reason why we do not find convergence to efficiency in the setting of this paper is that the firm's endogenous separation payoff (i.e., the payoff that the firm expects from going back to the labor market and offering the profit-maximizing contract to each new manager) is lower than the planner's endogenous separation payoff (i.e., the surplus that the planner expects by forcing the firm to go back to the labor market and offer the welfare-maximizing contract to each new manager). Indeed, the fact that each manager has private information about his own productivity at the time of contracting means that the firm cannot extract the full surplus from the relationship with each manager while inducing him to work efficiently. As explained above, the firm expects, at the time of hiring, to extract more surplus from the relationship with each incumbent as time goes by, with the flow payoff of the firm eventually converging to the flow total surplus that a benevolent planner would expect by retaining the same incumbent. The fact that the firm expects a lower payoff than the planner from going back to the labor market then implies that, eventually, the firm becomes excessively lenient in retaining its incumbents relative to what is efficient.[6]

This last result suggests that policy interventions aimed at inducing firms to sustain a higher turnover, for example, by offering them temporary tax incentives after a change in management or through the introduction of a mandatory retirement age for top employees, can, in principle, increase welfare.[7] Of course, such policies might be expected to encounter opposition on other grounds whose discussion is beyond the scope of this analysis.

Layout.—The rest of the paper is organized as follows. In the remainder of this section we briefly review the pertinent literature. Section II introduces the model. Section III characterizes the efficient contract. Section IV characterizes the firm's profit-maximizing contract and uses it to establish the key positive results. Section V compares the dynamics of retention decisions under the efficient contract with those under the profit-maximizing contract and establishes the key normative results. All proofs are in Appendix A.

Related literature.—The paper is related to various lines of research in the managerial compensation and turnover literature. A vast body of work documents how the threat of replacement plays an important role in incentivizing effort.[8] Recent contributions in this area include Clementi and Hopenhayn (2006), DeMarzo and Sannikov (2006), Tchistyi

[6] Note that this result also applies to a setting in which optimal effort is constant over time.

[7] See Lazear (1979) for alternative explanations for why mandatory retirement can be beneficial.

[8] Despite the vast attention that this property has received in the theoretical literature, the empirical evidence of the effect of turnover on incentives is mixed. See Jenter and Lewellen (2010) for a recent discussion and Gayle, Golan, and Miller (2008) for a recent empirical study of the relationship between promotion, turnover, and compensation in the market for executives.

(2006), Biais et al. (2007), DeMarzo and Fishman (2007), and He (2009). The reason why the threat of termination is essential in these papers is that the agent is protected by limited liability. This implies that incentives provided entirely through performance-based compensation need not be strong enough. The threat of termination is also crucial in the "efficiency wages" theory; in particular, see Shapiro and Stiglitz's (1984) seminal work. However, contrary to the literature cited above, in the efficiency wages theory, under the optimal contract, no worker shirks, and hence replacement does not occur in equilibrium.

Related to this line of research is also the work by Spear and Wang (2005), Sannikov (2008), and Wang (2011). These papers show how a risk-averse agent may be optimally induced to cease to exert effort and then retire once his promised continuation utility becomes either too high or too low, making it too costly for the firm to incentivize further effort.[9]

While not all the works cited above focus explicitly on turnover, they do offer implications for the dynamics of retention decisions. For example, Wang (2011) shows how a worker with a shorter tenure faces a higher probability of an involuntary layoff and a lower probability of voluntary retirement than a worker with a longer tenure. In a financial contracting setting, Clementi and Hopenhayn (2006) show how, on average, a borrower's promised continuation utility increases over time and how this requires an increase in the likelihood that the loan is rolled over. Similarly, Fong and Li (2010) find that the turnover rate eventually decreases in the duration of the employment relationship, but because contracts are relational, they also find that the turnover rate may initially increase. In the same spirit, Board (2011) finds that firms' retention decisions become inefficiently lenient after long tenure when they are governed by a relational contract.[10]

The above literature does not account for the possibility of changes in managerial productivity (equivalently, in the quality of the match between the manager and the firm). It therefore misses the possibility that turnover is driven by variations in managerial productivity in addition to concerns for incentivizing effort. Such a possibility has long been recognized as important by another body of the literature that dates back at

[9] Another paper in which dismissal helps to create incentives is Sen (1996). In this paper, the manager's private information is the productivity of the firm, which is assumed to be constant over time and independent of the manager who runs it. As in the current paper, commitments to replace the initial manager help to reduce informational rents. However, contrary to the current paper, there are no hidden actions and there is a single replacement decision. The analysis in Sen's paper thus does not permit one to study how the leniency of retention decisions evolves over time.

[10] A key difference between the result in Board (2011) and the one in the present paper is that, while inefficiency in his model originates in the firm's inability to commit to long-term contracts, which can be viewed as a form of "lack of good governance," in our model it is entirely due to asymmetric information.

least to Jovanovic (1979).[11] This paper considers an environment in which productivity (equivalently, the match quality) is constant over time but unknown to both the firm and the worker, who jointly learn it over time through the observation of realized output. Because of learning, turnover becomes less likely over time.[12] Our theory differs from Jovanovic's in a few respects. First, and important, we allow learning about match quality to be asymmetric between the workers and the firm, with the former possessing better information than the latter. Second, we explicitly model managerial effort and account for the fact that it must be incentivized. Third, we consider more general processes for the evolution of the match quality. These distinctions lead to important differences in the results. First, while in Jovanovic's model the leniency of turnover decisions originates from the accumulation of information over time, in our model turnover decisions become more lenient over time even when conditioning on the accuracy of available information (formally, even when the kernels, i.e., the transition probabilities, remain constant over time). Second, while in Jovanovic's model turnover decisions are always second-best efficient, in our model, turnover decisions are second-best inefficient and the inefficiency of such decisions typically increases over time.[13]

More recent papers in which turnover is also driven by variations in match quality include Acharya (1992), Mortensen and Pissarides (1994), Atkeson and Cole (2005), and McAdams (2011). Acharya (1992) studies how the market value of a firm changes after the announcement to replace a chief executive officer and how the probability of replacement is affected by the CEO's degree of risk aversion.[14] Mortensen and Pissarides (1994) show how the optimal turnover policy takes the form of a simple threshold policy, with the threshold being constant over time. Along with the assumption that productivity is drawn independently each time it changes and the fact that the revisions follow a Poisson process, this implies that the probability of terminating a relationship does not vary with tenure. In contrast, in a model of stochastic partnerships, McAdams (2011) finds that relationships become more stable over time because of a survivorship bias. Atkeson and Cole (2005) show how managers who delivered high performance in the past have a higher contin-

[11] Allgood and Farrell (2003) provide empirical support for the importance of variations in managerial productivity and, more generally, in match quality for turnover decisions.

[12] Also related is Holmström's (1999) career concerns model. While this paper does not characterize the optimal turnover policy, the evolution of career concerns has been recognized as a possible determinant for turnover; see, e.g., Mukherjee (2008).

[13] Inefficiencies originate in our theory from the combination of asymmetric information at the contracting stage with search frictions. Because neither the firms nor the managers can appropriate the entire surplus, contractual decisions are distorted relative to their second-best counterparts.

[14] Acharya (1992) also documents the possible optimality of permanently tenuring a CEO, a possibility that we also accommodate but show to never be optimal in our model.

uation utility and are then optimally rewarded with job stability. Because a longer tenure implies a higher probability of having delivered high performance in the past, their model also offers a possible explanation for why retention decisions may become more lenient over time.

An important distinction between our paper and the two bodies of the literature discussed above is that, in our theory, variations in match quality are anticipated but privately observed. As a result, a properly designed contract must not only incentivize effort but also provide managers with incentives for truthfully reporting to the board variations in match quality that call for adjustments in the compensation scheme and possibly for separation decisions. The importance of private information for turnover decisions has been recognized by another body of the literature that includes Levitt and Snyder (1997), Banks and Sundaram (1998), Eisfeldt and Rampini (2008), Gayle et al. (2008), Inderst and Mueller (2010), and Yang (forthcoming). Some of these papers show how asymmetric information may lead to a form of entrenchment, that is, to situations in which the agent remains in place (or the project continues) although the principal would prefer ex post to replace him (or discontinue the project). What is missing in this literature is an account of the possibility that the managers' private information may change over time and hence an analysis of how the leniency of optimal turnover decisions evolves with the managers' tenure in the firm.[15]

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

From a methodological viewpoint, the paper builds on recent developments in the theory of dynamic mechanism design with persistent shocks to the agents' private information and in particular on Pavan et al. (2012).[16] Among other things, that paper (i) establishes an envelope theorem for dynamic stochastic problems that is instrumental to the design of

[15] An exception is Gayle et al. (2008). The authors use a longitudinal data set to evaluate the importance of moral hazard and job experience in jointly determining promotion, turnover rates, and compensation and to study how the latter changes across the different layers of an organization. The focus of their analysis is, however, very different from ours.

[16] The literature on dynamic mechanism design goes back to the pioneering work of Baron and Besanko (1984) and Besanko (1985). More recent contributions include Courty and Li (2000), Battaglini (2005), Athey and Segal (2007), Eso and Szentes (2007), Board (2008), Gershkov and Moldovanu (2009*a*, 2009*b*, 2010*a*, 2010*b*, 2010*c*, 2012), Bergemann and Välimäki (2010), Board and Skrzypacz (2010), Dizdar, Gershkov, and Moldovanu

optimal dynamic mechanisms and (ii) shows how the dynamics of distortions is driven by the dynamics of the impulse responses of the future types to the initial ones. The current paper applies these insights and, more generally, the methodology of Pavan et al. to a managerial contracting environment. It also shows how the techniques in their paper must be adapted to accommodate moral hazard in a non-time-separable dynamic mechanism design setting. The core (and distinctive) contribution of the present paper is, however, in the predictions that the theory identifies for the joint dynamics of effort, retention, and compensation.

Also related is Garrett and Pavan (2011*a*). That work shares with the present paper the same managerial contracting framework. However, it completely abstracts from the possibility of replacement, which is the focus of the present paper. Instead, it investigates how the optimality of seniority-based schemes (i.e., schemes that provide managers with longer tenure with more high-powered incentives) is affected by the managers' degree of risk aversion.[17] In particular, that paper shows that, under risk neutrality and declining impulse responses, optimal effort increases, on average, with time. The same property holds in the present paper but is not essential for the dynamics of retention decisions. In fact, while we find it instructive to relate these dynamics to the ones for effort, neither our positive nor our normative results hinge on the property that effort, on average, increases with tenure: the same results hold if the firm is constrained to ask the same level of effort from the manager in all periods.[18]

Also obviously related is the entire literature on dynamic managerial compensation without replacement. This literature is too vast to be successfully summarized here. We refer the reader to Edmans and Gabaix (2009) for an overview. See also Edmans and Gabaix (2011) and Edmans et al. (2012) for recent contributions in which, as in Laffont and Tirole (1986) and in the current paper, the moral hazard problem is solved using techniques from the mechanism design literature. These works con-

(2011), Garrett (2011), Pai and Vohra (2011), and Said (forthcoming). For a survey of these papers see Bergemann and Said (2011).

The analysis in the current paper, as well as in Pavan et al. (2012), is in discrete time. Recent contributions in continuous time include Zhang (2009), Strulovici (2011), and Williams (2011). These works show how the solution to a class of dynamic adverse selection problems with persistent private information (but without replacement) can be obtained in a recursive way with the level and derivative of promised utility as state variables. In contrast, both the optimal and the efficient contracts in our paper are obtained through a fixed-point dynamic programming problem whose solution is not recursive, thus permitting us to show how effort, compensation, and retention decisions depend explicitly on the entire history of productivity shocks.

[17] While, for simplicity, the current paper does not account for the possibility that the managers are risk averse, we expect our key predictions to remain true for a low degree of risk aversion.

[18] For example, dynamics of retention decisions qualitatively similar to the ones in this paper arise in an environment in which effort can take only negative values, say $e \in [-K, 0]$, and in which $e = 0$ is interpreted as "no stealing" and is optimally sustained at all periods, as in DeMarzo and Fishman (2007).

sider a setting in which (i) there is no turnover, (ii) managers possess no private information at the time of contracting, and (iii) it is optimal to induce a constant level of effort over time. Relaxing points i and ii is essential to our results. As explained above, endogenizing effort is also important for our predictions about the joint dynamics of effort, retention, and compensation but is not essential to the key properties identified in this paper.

II. Model

Players.—A principal (the board of directors, acting on behalf of the shareholders of the firm) is in charge of designing a new employment contract to govern the firm's interaction with its managers.[19] The firm is expected to operate for infinitely many periods, and each manager is expected to live as long as the firm. There are infinitely many managers. All managers are ex ante identical, meaning that they have the same preferences and that their productivity (to be interpreted as their ability to generate cash flows for the firm) is drawn independently from the same distribution and is expected to evolve over time according to the same Markov process described below.

Stochastic process.—The process governing the evolution of each manager's productivity is assumed to be independent of calendar time and exogenous to the firm's decisions. This process has two components: the distribution from which each manager's initial productivity is drawn and the family of conditional distributions describing how productivity evolves upon joining the firm.

For each $t \geq 1$, let θ_t denote a manager's productivity in the tth period of employment. Each manager's productivity during the first period of employment coincides with his productivity prior to joining the firm. This productivity is drawn from the absolutely continuous distribution F_1 with support $\Theta = (\underline{\theta}, \bar{\theta}) \subset \mathbb{R}$ and density function f_1. The distribution F_1 is meant to capture the distribution of managerial talent in the population.

For all $t > 1$, θ_t is drawn from the cumulative distribution function $F(\cdot|\theta_{t-1})$ with support Θ.[20] We assume that the function F is continuously

[19] As anticipated above, the focus of the analysis is on the contracts offered by a representative firm for given contracts offered by all other competing firms (equivalently, for given managers' outside options). However, the profit-maximizing and efficient contracts characterized below are also equilibrium and welfare-maximizing contracts in a setting in which unemployed managers are randomly matched with many (ex ante identical) firms. Indeed, as will become clear, as long as the number of potential managers is large compared to the number of competing firms, so that the matching probabilities remain independent of the contracts selected, the managers' outside options (i.e., their payoff after separation occurs) have an effect on the level of compensation but not on the profit-maximizing and efficient effort and retention policies.

[20] The process is thus *time autonomous*: the kernels are independent of the length of the employment relationship so that $F_t(\cdot|\cdot) = F(\cdot|\cdot)$ all $t > 1$. Each kernel has support on the

differentiable over Θ^2 and denote by $f(\theta_t|\theta_{t-1}) \equiv \partial F(\theta_t|\theta_{t-1})/\partial\theta_t$ the density of the cumulative distribution $F(\cdot|\theta_{t-1})$. We assume that, for any θ_t, $\theta_{t-1} \in \Theta$, $-f(\theta_t|\theta_{t-1}) \leq \partial F(\theta_t|\theta_{t-1})/\partial\theta_{t-1} \leq 0$. This guarantees (i) that the conditional distributions can be ranked according to first-order stochastic dominance and (ii) that the impulse responses (which are defined below and which capture the process's degree of persistence) are uniformly bounded.[21]

Given F_1 and the family $F \equiv \langle F(\cdot|\theta)\rangle_{\theta\in\Theta}$ of conditional distributions, we then define the *impulse responses* of future productivity to earlier productivity as follows (the definition here parallels that in Pavan et al. [2012]). Let $\tilde{\varepsilon}$ be a random variable uniformly distributed over $\mathcal{E} = [0, 1]$ and note that, for any $\theta \in \Theta$, the random variable $z(\theta, \tilde{\varepsilon}) \equiv F^{-1}(\tilde{\varepsilon}|\theta)$ is distributed according to $F(\cdot|\theta)$ by the integral transform probability theorem. For any $\tau \in \mathbb{N}$ then, let $Z_\tau : \Theta \times \mathcal{E}^\tau \to \Theta$ be the function defined inductively as follows: $Z_1(\theta, \varepsilon) \equiv z(\theta, \varepsilon)$, $Z_2(\theta, \varepsilon_1, \varepsilon_2) = z(Z_1(\theta, \varepsilon_1), \varepsilon_2)$, and so forth.[22] For any s and t, $s < t$, and any continuation history $\theta^t_{\geq s} \equiv (\theta_s, \ldots, \theta_t)$, the impulse response of θ_t to θ_s is then defined by

$$J^t_s(\theta^t_{\geq s}) \equiv \frac{\partial Z_{t-s}(\theta_s,\ \varepsilon^{t-s}(\theta^t_{\geq s}))}{\partial\theta_s},$$

where $\varepsilon^{t-s}(\theta^t_{\geq s})$ denotes the unique sequence of shocks that, starting from θ_s, leads to the continuation history $\theta^t_{\geq s}$. These impulse response functions are the nonlinear analogues of the familiar constant linear impulse responses for autoregressive processes. For example, in the case of an AR(1) process with persistence parameter γ, the impulse response of θ_t to θ_s is simply given by the scalar $J^t_s = \gamma^{t-s}$. More generally, the impulse response $J^t_s(\theta^t_{\geq s})$ captures the effect of an infinitesimal variation of θ_s on θ_t, holding constant the shocks $\varepsilon^{t-s}(\theta^t_{\geq s})$. As shown below, these functions play a key role in determining the dynamics of profit-maximizing effort

same interval Θ that defines the support of the period 1 distribution F_1. Both of these assumptions, as well as many of the technical conditions below, are stronger than needed for our results but simplify the exposition. See the working paper version of this paper (Garrett and Pavan 2011*b*) for how to accommodate non-time-autonomous processes with shifting supports and Pavan et al. (2012) for how to relax some of the technical conditions. On the other hand, allowing for more than two periods is essential to our results about the dynamics of retention decisions. Allowing for more than two productivity levels is also essential. In fact, one can easily verify that, with two productivity levels, the optimal retention policy takes one of the following three forms: (i) either the manager is never replaced, irrespective of the evolution of his productivity; (ii) or he is retained if and only if his initial productivity was high; (iii) or he is fired as soon as his productivity turns low. In each case, 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.

[21] The lower bound on $\partial F(\theta_t|\theta_{t-1})/\partial\theta_{t-1}$ is equivalent to assuming that, for any $\theta_{t-1} \in \Theta$, any $x \in \mathbb{R}$, $1 - F(\theta_{t-1} + x|\theta_{t-1})$ is nonincreasing in θ_{t-1}. That is, the probability that a manager's productivity in period t exceeds the one in the previous period by more than x is nonincreasing in the previous period's productivity.

[22] Throughout the entire paper, we will use superscripts to denote sequences of variables.

and turnover policies. Throughout, we will maintain the assumption that types evolve independently across managers.

Effort, cash flows, and payoffs.—After learning his period t productivity θ_t, the manager currently employed by the firm must choose an effort level $e_t \in E = \mathbb{R}$.[23] The firm's per-period cash flows, gross of the manager's compensation, are given by

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

where ν_t is transitory noise. The shocks ν_t are independent and identically distributed over time, independent across managers, and drawn from the distribution Φ, with expectation $\mathbb{E}[\tilde{\nu}_t] = 0$. The sequences of productivities θ^t and effort choices $e^t \equiv (e_1, \ldots, e_t) \in E^t$ are the manager's private information. In contrast, the history of cash flows $\pi^t \equiv (\pi_1, \ldots, \pi_t) \in \mathbb{R}^t$ generated by each manager is verifiable and can be used as a basis for compensation.

By choosing effort $e \in E$ in period t, the manager suffers a disutility $\psi(e) \geq 0$, where $\psi(\cdot)$ is a differentiable and Lipschitz continuous function with $\psi(0) = 0$. As in Laffont and Tirole (1986), we assume that there exists a scalar $\bar{e} > 0$ such that ψ is thrice continuously differentiable over $(0, \bar{e})$ with $\psi'(e), \psi''(e) > 0$, and $\psi'''(e) \geq 0$ for all $e \in (0, \bar{e})$ and that $\psi'(e) > 1$ for all $e > \bar{e}$.[24] These last properties guarantee that both the efficient and the profit-maximizing effort levels are interior while ensuring that the manager's payoff is equi-Lipschitz continuous in effort. The latter property permits us to conveniently express the value function through a differentiable envelope formula (more below).[25]

Denoting by c_t the compensation that the manager receives in period t (equivalently, his period t consumption), the manager's preferences over (lotteries over) streams of consumption levels $c \equiv (c_1, c_2, \ldots)$ and streams of effort choices $e \equiv (e_1, e_2, \ldots)$ are described by an expected utility function with (Bernoulli) utility given by

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

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 managerial compensation. Formally, let π_{it} and c_{it} denote, respectively, the cash flow generated and the compensation received by the ith manager employed by the firm in his tth period of employment. Then, let T_i denote the num-

[23] The assumption that effort takes on any real value is made only for simplicity.

[24] Note that these conditions are satisfied, e.g., 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 $e > \bar{e}$.

[25] None of the results hinge on the value of $\bar{e}$. Indeed, the firm's payoff is invariant to $\bar{e}$ (holding constant ψ over the interval $\{e : 0 \leq \psi'(e) \leq 1\}$).

ber of periods for which manager i works for the firm. The contribution of manager i to the firm's payoff, evaluated at the time manager i is hired, is given by

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

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

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

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

Timing and labor market.—The firm's interaction with the labor market unfolds as follows. Each manager learns his initial productivity θ_1 prior to being matched with the firm. After being matched, the manager is offered a menu of contracts described in detail below. While the firm can perfectly commit to the contracts it offers, each manager is free to leave the firm at each point in time. After leaving the firm, the manager receives a continuation payoff equal to $U^o \geq 0$.[26]

We assume (i) that it is never optimal for the firm to operate without a manager being in control, (ii) that it is too costly to sample another manager before separating from the incumbent, and (iii) that all replacement decisions must be planned at least one period in advance. These assumptions capture (in a reduced form) various frictions in the recruiting process that prevent firms from sampling until they find a manager of the highest possible productivity, which is unrealistic and would make the analysis uninteresting.[27]

[26] That the outside option is invariant to the manager's productivity is a simplification. All our results extend qualitatively to a setting in which the outside option is type dependent as long as the derivative of the outside option $U^o(\theta_t)$ with respect to current productivity is sufficiently small that the single-crossing conditions of Sec. IV are preserved. This is the case, e.g., when (i) the discount factor is not very high and/or (ii) it takes a long time for a manager to find a new job. Also note that, from the perspective of the firm under examination, this outside option is exogenous. However, in a richer setting with multiple identical firms and exogenous matching probabilities, U^o will coincide with the equilibrium continuation payoff that each manager expects from going back to the labor market and being randomly matched (possibly after an unemployment phase) with another firm. In such an environment, each manager's outside option is both time and type invariant (and equal to zero) if there are infinitely more managers than firms.

[27] The assumption of random matching is also quite standard in the labor/matching literature (see, e.g., Jovanovic 1979). In our setting, it implies that there is no direct competition among managers for employment contracts. This distinguishes our environment from an auction-like setting in which, in each period, the principal consults simultaneously with multiple managers and then chooses which one to hire/retain.

After signing one of the contracts, the manager privately chooses effort e_1. Nature then draws ν_1 from the distribution Φ and the firm's (gross) cash flows π_1 are determined according to (1). After observing the cash flows π_1, the firm pays the manager a compensation c_1, which may depend on the specific contract selected by the manager and on the verifiable cash flow π_1. On the basis of the specific contract selected at the time of contracting and the observed cash flow, the manager is then either retained or dismissed at the end of the period.[28] If the manager is retained, his second-period productivity is then drawn from the distribution $F(\cdot|\theta_1)$. After privately learning θ_2, at the beginning of the second period of employment, the manager then decides whether or not to leave the firm. If he leaves, he obtains the continuation payoff U^o. If he stays, he is then offered the possibility of modifying the terms of the contract that pertain to future compensation and retention decisions within limits specified by the contract signed in the first period (as it will become clear in a moment, these adjustments are formally equivalent to reporting the new productivity θ_2). After these adjustments are made, the manager privately chooses effort e_2, cash flows π_2 are realized, and the manager is then paid a compensation c_2 as specified by the original contract along with the adjustments made at the beginning of the second period (clearly, the compensation c_2 may also depend on the entire history of observed cash flows $\pi^2 = (\pi_1, \pi_2)$). Given the contract initially signed, the adjustments made in period 2, and the observed cash flows π^2, the manager is then either retained into the next period or dismissed at the end of the period.

The entire sequence of events described above repeats itself over time until the firm separates from the manager or the manager unilaterally decides to leave the firm. After separation occurs, at the beginning of the subsequent period, the firm goes back to the labor market and is randomly matched with a new manager whose initial productivity θ_1 is drawn from the same stationary distribution F_1 from which the incumbent's initial productivity was drawn. The relationship between any newly sampled manager and the firm then unfolds in the same way as described above for the incumbent.

The employment relationship as a dynamic mechanism.—Because all managers are ex ante identical, time is infinite, and types evolve independently across managers, the firm offers the same menu of contracts to each manager it is matched with. Under any such contract, the compensation that the firm pays to the manager (as well as the retention decisions) may de-

[28] That retention decisions are specified explicitly in the contract simplifies the exposition but is not essential. For example, by committing to pay a sufficiently low compensation after all histories that are supposed to lead to separation, the firm can always implement the desired retention policy by delegating to the managers the choice of whether or not to stay in the relationship. It will become clear from the analysis below that, while both the optimal and the efficient retention policies are unique, there are many ways these policies can be implemented (see, e.g., Yermack [2006] for a description of the most popular termination clauses and "golden handshakes" practices).

pend on the cash flows produced by the manager as well as on messages sent by the manager over time (as explained above, the role of these messages is to permit the firm to respond to variations in productivity). However, both compensation and retention decisions are independent of both the calendar time at which the manager was hired and the history of messages sent and cash flows generated by other managers. Hereafter, we will thus maintain the notation that t denotes the number of periods that a representative manager has been working for the firm and not the calendar time.

Furthermore, because the firm can commit, one can conveniently describe the firm's contract as a direct revelation mechanism. This specifies, for each period t, a recommended effort choice, the contingent compensation, and a retention decision.

In principle, both the level of effort recommended and the retention decision may depend on the history of reported productivities and on the history of cash flow realizations. However, it can be shown that, under both the efficient and profit-maximizing contracts, the optimal effort and retention decisions depend only on reported productivities θ^t.[29] The reason is that any type of manager, by adjusting his effort level, can generate the same cash flow distribution as any other type, regardless of the other type's effort level and regardless of the noise distribution (in particular, even if the noise is absent). Cash flows are thus a very weak signal of productivity—which is the only serially correlated state variable—and hence play no prominent role in retention and future effort decisions, which are decisions about productivity.[30] On the other hand, because the effort decisions are "hidden actions" (i.e., because of moral hazard), it is essential that the total compensation be allowed to depend both on the reported productivities θ^t and on past and current cash flows π^t.

Hereafter, we will thus model the employment relationship induced by the profit-maximizing and the efficient contracts as a direct revelation mechanism $\Omega \equiv \langle \xi, x, \kappa \rangle$. This consists of a sequence of functions $\xi \equiv (\xi_t : \Theta^t \to E)_{t=1}^{\infty}$, $x \equiv (x_t : \Theta^t \times \mathbb{R}^t \to \mathbb{R})_{t=1}^{\infty}$, and $\kappa \equiv (\kappa_t : \Theta^t \to \{0, 1\})_{t=1}^{\infty}$ such that

- $\xi_t(\theta^t)$ is the recommended period t effort;
- $x_t(\theta^t, \pi^t)$ is the compensation paid at the end of period t;

[29] A formal proof for this result can be found in online App. B.

[30] Note that this result would not hold if the manager were risk averse. The reason is that conditioning retention and effort decisions on past and current cash flows can help reduce the firm's cost of shielding a risk-averse manager from risk. The result would also not be true if the manager were cash constrained, in which case committing to fire him after a poor performance may be necessary to incentivize his effort.

- $\kappa_t(\theta^t)$ is the retention decision for period t, with $\kappa_t(\theta^t) = 1$ if the manager is to be retained, which means that he is granted the possibility of working for the firm also in period $t + 1$, regardless of his period $t + 1$ productivity θ_{t+1},[31] and $\kappa_t(\theta^t) = 0$ if either (i) he is dismissed at the end of period t or (ii) he was dismissed in previous periods; that is, $\kappa_t(\theta^t) = 0$ implies $\kappa_s(\theta^s) = 0$ for all $s > t$, all θ^s.[32] Given any sequence θ^∞, we then denote by $\tau(\theta^\infty) \equiv \min\{t : \kappa_t(\theta^t) = 0\}$ the corresponding *length* of the employment relationship.

In each period t, given the previous reports $\hat{\theta}^{t-1}$ and cash flow realizations π^{t-1}, the employment relationship unfolds as follows:

- After learning his period t productivity $\theta_t \in \Theta_t$ and upon deciding to stay in the relationship, the manager sends a report $\hat{\theta}_t \in \Theta_t$.
- The mechanism then prescribes effort $\xi_t(\hat{\theta}^{t-1}, \hat{\theta}_t)$ and specifies a reward scheme $x_t(\hat{\theta}^{t-1}, \hat{\theta}_t, \pi^{t-1}, \cdot) : \mathbb{R} \to \mathbb{R}$ along with a retention decision $\kappa_t(\hat{\theta}^{t-1}, \hat{\theta}_t)$.
- The manager then chooses effort e_t.
- After observing the realized cash flows $\pi_t = e_t + \theta_t + \nu_t$, the manager is paid $x_t(\hat{\theta}^{t-1}, \hat{\theta}_t, \pi^{t-1}, \pi_t)$ 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 manager, and (ii) after any truthful and obedient history, the manager finds it optimal to stay in the relationship whenever offered the possibility of doing so (i.e., the manager never finds it optimal to leave the firm when he has the option to stay). In the language of dynamic mechanism design, the first property means that the mechanism is *incentive compatible* and the second property means that it is *sequentially individually rational.*

Remark.—While we are not imposing limited liability (or cash) constraints on the principal's problem, the effort and retention policies that we characterize below turn out to be implementable with nonnegative payments for reasonable parameter specifications (see corollary 1 below).

[31] Recall that separation decisions must be planned one period in advance and that it is too costly to go back to the labor market and consult another manager before separating from the incumbent. Along with the assumption that it is never desirable to operate the firm without a manager, these assumptions imply that a manager who is retained at the end of period t will never be dismissed at the beginning of period $t + 1$, irrespective of his period $t + 1$ productivity.

[32] For expositional convenience, we allow the policies ξ_t, x_t, and κ_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.

III. The Efficient Contract

We begin by describing the effort and turnover policies ξ^E and κ^E that maximize ex ante welfare, defined to be the sum of a representative manager's expected payoff and of the firm's expected profits (the "efficient" policies). Although we are clearly interested in characterizing these policies for the same environment as described above, it turns out that these policies coincide with the ones that maximize ex ante welfare in an environment with symmetric information, in which the managers' productivities and effort choices are observable and verifiable. In turn, because all players' payoffs are linear in payments, these policies also coincide with the ones that the firm would choose under symmetric information to maximize expected profits. For simplicity, in this section, we thus assume that information is symmetric and then show in Section V—proposition 7—that the efficient policies under symmetric information remain implementable also under asymmetric information.

The efficient effort policy is very simple: Because all players are risk neutral and because each manager's productivity has no effect on the marginal cost or the marginal benefit of effort, the efficient effort level e^E is independent of the history of realized productivities and is implicitly defined by the first-order condition $\psi'(e^E) = 1$.

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 or the productivities of its future hires, this problem involves a trade-off in each period between experimenting with a new manager and continuing experimenting with the incumbent. Denote by $\mathcal{B}^E$ the set of all bounded functions from Θ 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 $\theta \in \Theta$ and irrespective of t, gives the firm's expected continuation payoff when the incumbent manager's productivity is θ.[33] Clearly, the value $W^E(\theta)$ takes into account the possibility of replacing the manager in the future. As we show in Appendix A, the function W^E is the unique fixed point to the mapping $T_E : \mathcal{B}^E \to \mathcal{B}^E$ defined, for all $W \in \mathcal{B}^E$, all $\theta \in \Theta$, by[34]

$$T_E W(\theta) = \theta + e^E - \psi(e^E) - (1-\delta)U^o + \delta \max\{\mathbb{E}_{\tilde{\theta}|\theta}[W(\tilde{\theta})];\ \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1)]\}.$$

The efficient contract can then be described as follows.

[33] Note that if the process were not autonomous, the efficient retention decision would obviously depend also on the length t of the employment relationship. See the working paper version of this paper (Garrett and Pavan 2011*b*) for how the result in the next proposition must be adapted to accommodate nonautonomous processes.

[34] The expectations $\mathbb{E}_{\tilde{\theta}|\theta}[W(\tilde{\theta})]$ and $\mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1)]$ are, respectively, under the measures $F(\cdot|\theta)$ and $F_1(\cdot)$: recall that, under the simplifying assumption that the process is autonomous, for any $t > 1$, any $\theta \in \Theta$, $F_t(\cdot|\theta) = F(\cdot|\theta)$.

PROPOSITION 1. The efficient effort and turnover policies satisfy the following properties:[35] (i) For all t, all $\theta^t \in \Theta^t$, $\xi_t^E(\theta^t) = e^E$, with e^E implicitly defined by $\psi'(e^E) = 1$. (ii) Conditional on being employed in period t, the manager is retained at the end of period t if and only if $\theta_t \geq \theta^E$, where

$$\theta^E = \inf\{\theta \in \Theta : \mathbb{E}_{\tilde{\theta}|\theta}[W^E(\tilde{\theta})] \geq \mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1)]\}.$$

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 $T_E : \mathcal{B}^E \to \mathcal{B}^E$ defined above. It then shows that this function is indeed the value function for the problem described above. Finally, it establishes that the function W^E is nondecreasing. These properties, together with the assumptions that the process is Markov, autonomous, and with kernels that can be ranked according to first-order stochastic dominance, imply that turnover decisions must be made according to the cutoff rule given in the proposition.

IV. The Profit-Maximizing Contract

We now turn to the contract that maximizes the firm's expected profits in a setting in which neither the managers' productivities nor their effort choices are observable. As anticipated above, what prevents the firm from appropriating the entire surplus (equivalently, from "selling out" the project to the managers) is the fact that, both at the initial contracting stage and at any subsequent period, each manager is privately informed about his productivity. To extract some of the surplus from the most productive types, the firm must then introduce distortions in effort and retention decisions, which require retaining ownership of the project.

We start by showing that, in any incentive-compatible mechanism $\Omega \equiv \langle \xi, x, \kappa \rangle$, each type's intertemporal expected payoff under a truthful and obedient strategy $V^\Omega(\theta_1)$ must satisfy

$$V^\Omega(\theta_1) = V^\Omega(\underline{\theta}) + \int_{\underline{\theta}}^{\theta_1} \mathbb{E}_{\tilde{\theta}^\infty_{>1}|s}\left[\sum_{t=1}^{\tau(s,\tilde{\theta}^\infty_{>1})} \delta^{t-1} J_1^t(s, \tilde{\theta}^t_{>1}) \psi'(\xi_t(s, \tilde{\theta}^t_{>1}))\right] ds. \quad (4)$$

The derivation of this formula follows from arguments similar to those in Pavan et al. (2012), adapted to the environment under examination here. To establish (4), consider the following fictitious environment in which the manager can misrepresent his type but is then "forced" to

[35] The efficient policies are "essentially unique," i.e., unique up to a zero-measure set of histories.

choose effort so as to hide his lies by inducing the same distribution of cash flows as if his reported type coincided with the true one. This is to say that, at any period t, given the history of reports $\hat{\theta}^t$ and the true current productivity θ_t, the manager must choose effort

$$e_t^{\#}(\theta_t; \hat{\theta}^t) = \hat{\theta}_t + \xi_t(\hat{\theta}^t) - \theta_t \tag{5}$$

so that the distribution of the period t cash flows is the same as when the manager's true period t productivity is $\hat{\theta}_t$ and the manager follows the recommended effort choice $\xi_t(\hat{\theta}^t)$.

Clearly, if the mechanism Ω is incentive compatible and sequentially individually rational in the original environment in which the manager is free to choose his effort after misreporting his type, it must also be in this fictitious one, where he is forced to choose effort according to (5). This allows us to focus on a necessary condition for the optimality of truthful reporting by the manager in the fictitious environment, which remains necessary for such behavior in the original one.

Fix an arbitrary sequence of reports $\hat{\theta}^\infty$ and an arbitrary sequence of true productivities θ^∞. Let $C(\hat{\theta}^\infty)$ denote the present value of the stream of payments that the manager expects to receive from the principal when the sequence of reported productivities is $\hat{\theta}^\infty$ and, in each period, he chooses effort according to (5).[36] For any $(\theta^\infty, \hat{\theta}^\infty)$, the manager's expected payoff in this fictitious environment is given by

$$\begin{aligned}\mathcal{U}(\theta^\infty, \hat{\theta}^\infty) \equiv C(\hat{\theta}^\infty) &- \sum_{t=1}^{\infty} \delta^{t-1}\kappa_{t-1}(\hat{\theta}^{t-1})\psi(\hat{\theta}_t + \xi_t(\hat{\theta}^t) - \theta_t) \\ &+ \sum_{t=1}^{\infty} \delta^{t-1}[1 - \kappa_{t-1}(\hat{\theta}^{t-1})](1-\delta)U^o.\end{aligned}$$

The assumption that ψ is differentiable and Lipschitz continuous implies that $\mathcal{U}$ is totally differentiable in θ^t, any t, and equi-Lipschitz continuous in θ^∞ in the norm

$$||\theta^\infty|| \equiv \sum_{t=1}^{\infty} \delta^t |\theta_t|.$$

Together with the fact that $||\theta^\infty||$ is finite (which is implied by the assumption that Θ is bounded) and that the impulse responses $J_s^t(\theta^t)$ are uniformly bounded, this means that the dynamic envelope theorem of Pavan et al. (2012, proposition 3) applies to this environment. Hence, a necessary condition for truthful reporting to be optimal for the manager

[36] Note that, by construction, C does not depend on the true productivities θ^∞. Also note that the expectation here is over the transitory noise v^∞.

in this fictitious environment (and by implication also in the original one) is that the value function $V^{\Omega}(\theta_1)$ associated to the problem that involves choosing the reports and then selecting effort according to (5) is Lipschitz continuous and, at each point of differentiability, satisfies

$$\frac{dV^{\Omega}(\theta_1)}{d\theta_1} = \mathbb{E}_{\tilde{\theta}^{\infty}_{>1}|\theta_1}\left[\sum_{t=1}^{\infty}\delta^{t-1}J_1^t(\theta_1, \tilde{\theta}^t_{>1})\frac{\partial\mathcal{U}(\tilde{\theta}^{\infty}, \tilde{\theta}^{\infty})}{\partial\theta_t}\right],$$

where $\partial\mathcal{U}(\theta^{\infty}, \theta^{\infty})/\partial\theta_t$ denotes the partial derivative of $\mathcal{U}(\theta^{\infty}, \theta^{\infty})$ with respect to the true (rather than the reported) type θ_t. The result then follows from the fact that

$$\frac{\partial\mathcal{U}(\theta^{\infty}, \theta^{\infty})}{\partial\theta_t} = \kappa_{t-1}(\theta^{t-1})\psi'(\xi_t(\theta^t))$$

and the definition of the stopping time $\tau(\theta^{\infty}) \equiv \min\{t : \kappa_t(\theta^t) = 0\}$.

The formula in (4) confirms the intuition that the expected surplus that the principal must leave to each period 1 type is determined by the dynamics of effort and retention decisions under the contracts offered to the less productive types. As anticipated in the introduction, the reason is that those managers who are most productive at the contracting stage expect to be able to obtain a "rent" when mimicking the less productive types. This rent originates from the possibility of generating the same cash flows as the less productive types by working less, thus economizing on the disutility of effort. The amount of effort they expect to save must, however, take into account the fact that their own productivity, as well as that of the types they are mimicking, will change over time. This is done by weighting the amount of effort saved in all subsequent periods by the impulse response functions J_1^t, which, as explained above, control for how the effect of the initial productivity on future productivity evolves over time.

Now let

$$\eta(\theta_1) \equiv \frac{1 - F_1(\theta_1)}{f_1(\theta_1)}$$

denote the inverse hazard rate of the first-period distribution. Then (4) gives the following useful result (the proof follows from the arguments above).

Proposition 2. In any incentive-compatible and sequentially individually rational mechanism $\Omega \equiv \langle\xi, x, \kappa\rangle$, the firm's expected profits from each manager it hires are given by

$$\mathbb{E}_{\tilde{\theta}^\infty,\tilde{\nu}^\infty}\left[\sum_{t=1}^{\tau(\tilde{\theta}^\infty)}\delta^{t-1}[\tilde{\theta}_t+\xi_t(\tilde{\theta}^t)+\tilde{\nu}_t-\psi(\xi_t(\tilde{\theta}^t)) \right.$$
$$\left. -\ \eta(\tilde{\theta}_1)J_1^t(\tilde{\theta}^t)\psi'(\xi_t(\tilde{\theta}^t))-(1-\delta)U^o]\right]+U^o-V^\Omega(\underline{\theta}), \tag{6}$$

where $V^\Omega(\underline{\theta})\geq U^o$ denotes the expected payoff of the lowest period 1 type.

The formula in (6) is the dynamic analogue of the familiar virtual surplus formula for static adverse selection settings. It expresses the firm's expected profits as the discounted expected total surplus generated by the relationship, net of terms that control for the surplus that the firm must leave to the manager to induce him to participate in the mechanism and to truthfully reveal his private information.

Equipped with the aforementioned representation, we now consider a "relaxed program" that involves choosing the policies $(\xi_t(\cdot),\ \kappa_t(\cdot))_{t=1}^\infty$ so as to maximize the expected total payoff of the firm, taking the contribution of each manager to be (6) (note that this incorporates only the local incentive constraints) and subject to the participation constraints of the lowest period 1 types $V^\Omega(\underline{\theta})\geq U^o$.

Below, we first characterize the policies $(\xi_t^*(\cdot),\ \kappa_t^*(\cdot))_{t=1}^\infty$ that solve the relaxed program. We then provide sufficient conditions for the existence of a compensation scheme x^* such that the mechanism $\Omega\equiv\langle\xi^*,x^*,\kappa^*\rangle$ is incentive compatible and sequentially individually rational (and hence profit maximizing for the firm).

Let $A=\cup_{t=1}^\infty\Theta^t$ and denote by $\mathcal{B}$ the set of bounded functions from A to $\mathbb{R}$. For any effort policy ξ, let W_ξ^* denote the unique fixed point to the mapping $T(\xi):\mathcal{B}\to\mathcal{B}$ defined, for all $W\in\mathcal{B}$, all t, all θ^t, by

$$T(\xi)W(\theta^t)\equiv\xi_t(\theta^t)+\theta_t-\psi(\xi_t(\theta^t))-\eta(\theta_1)J_1^t(\theta^t)\psi'(\xi_t(\theta^t))$$
$$-\ (1-\delta)U^o+\delta\max\{\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W(\tilde{\theta}^{t+1})],\ \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1)]\}. \tag{7}$$

Proposition 3. Let ξ^* be the effort policy implicitly defined, for all t, all $\theta^t\in\Theta^t$, by[37]

$$\psi'(\xi_t^*(\theta^t))=1-\eta(\theta_1)J_1^t(\theta^t)\psi''(\xi_t^*(\theta^t)). \tag{8}$$

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

and (suppressing the dependence on ξ^* to ease the exposition) let W^* be the unique fixed point to the mapping $T(\xi^*)$ defined by (7). Let κ^* denote the retention policy such that, for any t and any $\theta^t \in \Theta^t$, conditional on the manager being employed in period t, he is retained at the end of period t if and only if $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] \geq \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$. The pair of policies (ξ^*, κ^*) solves the firm's relaxed program.

The effort and turnover policies that solve the relaxed program are thus the "virtual analogues" of the policies ξ^E and κ^E that maximize efficiency, as given in proposition 1. Note that, in each period t and for each history $\theta^t \in \Theta^t$, 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 1's perspective (i.e., at the time the managers are hired). The fact that both the firm's and the managers' preferences are additively separable over time implies that this trade-off is unaffected by the possibility that the firm replaces the managers. Furthermore, because each type θ_1's rent $V^\Omega(\theta_1)$ is increasing in the effort $\xi_t(\theta_1', \theta_{>1}^t)$ that the firm asks of each less productive type $\theta_1' < \theta_1$ in each period $t \geq 1$, the optimal effort policy is downward distorted relative to its efficient counterpart ξ^E, as in Laffont and Tirole's (1986) static model.

More interestingly, note that, fixing the initial type θ_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 1 productivity on subsequent productivity, capture how the persistence of the managers' initial private information evolves over time. Because such persistence is what makes more productive (period 1) types expect larger surplus in subsequent periods than initially less productive types, the dynamics of the impulse responses J_1^t are what determine the dynamics of effort decisions ξ_t^*.

Next, consider the turnover policy. The characterization of the profit-maximizing policy κ^* parallels the one for the efficient policy κ^E in proposition 1. The proof in Appendix A first establishes that the (unique) fixed point W^* to the mapping $T(\xi^*)$ given by (7) coincides with the value function associated with the problem that involves choosing the turnover policy so as to maximize the expected total virtual surplus (given for each manager by [6]) taking as given the profit-maximizing effort policy ξ^*. It then uses W^* to derive the optimal retention policy.

For any t, any $\theta^t \in \Theta^t$, $W^*(\theta^t)$ gives the firm's expected continuation profits (under all its future hires) when the incumbent manager has worked already for $t - 1$ periods and 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 $W^*(\theta^t)$ in general depends on the entire history of productivities θ^t as opposed to only the current productivity θ_t. There are two reasons. First, as shown above, the profit-maximizing effort policy typically depends on the entire history θ^t. Second, even if effort were exogenously fixed at a constant level, because productivity is serially correlated, conditioning the current retention decision on past productivity reports in addition to the current report is helpful in inducing the manager to have been truthful at the time he made those past reports.

The profit-maximizing turnover policy can then be determined straightforwardly from the value function W^*: each incumbent manager is replaced whenever the expected value $\mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$ of starting a relationship with a new manager of unknown productivity exceeds the expected value $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})]$ of continuing the relationship with the incumbent. Once again, these values are calculated from the perspective of the time at which the incumbent is hired and take into account the optimality of future effort and retention decisions.

Having characterized the policies that solve the relaxed program, we now turn to sufficient conditions that guarantee that such policies are indeed implemented under any optimal contract for the firm—in other words, solve the firm's full program (recall that [6] incorporates only local incentive-compatibility conditions, as implied by the envelope formula [4]).

We establish the result by showing existence of a compensation scheme x^* that implements the policies (ξ^*, κ^*) at minimal cost for the firm. In particular, given the mechanism $\Omega^* = (\xi^*, x^*, \kappa^*)$, the following properties hold true: (i) after any history $h_t = (\theta^t, \hat{\theta}^{t-1}, e^{t-1}, \pi^{t-1})$ such that $\kappa^*_{t-1}(\hat{\theta}^{t-1}) = 1$, each manager prefers to follow a truthful and obedient strategy in the entire continuation game that starts in period t with history h_t rather than any other strategy; (ii) the lowest period 1 type's expected payoff $V^{\Omega}(\underline{\theta})$ from following a truthful and obedient strategy in the entire game is exactly equal to his outside option U^o; and (iii) after any history $h_t = (\theta^t, \hat{\theta}^{t-1}, e^{t-1}, \pi^{t-1})$ such that $\kappa^*_{t-1}(\hat{\theta}^{t-1}) = 1$, each manager's continuation payoff under a truthful and obedient strategy remains at least as high as his outside option U^o. That the mechanism Ω^* is optimal for the firm then follows from the fact that the mechanism is incentive compatible and sequentially individually rational, along with the results in propositions 2 and 3.

Proposition 4. Suppose that the policies (ξ^*, κ^*) defined in proposition 3 satisfy the following *single-crossing conditions* for all $t \geq 1$, all $\theta_t, \hat{\theta}_t \in \Theta_t$, all $\hat{\theta}^{t-1} \in \Theta^{t-1}$ such that $\kappa_{t-1}(\hat{\theta}^{t-1}) = 1$:

$$\begin{aligned}\mathbb{E}_{\tilde{\theta}^{\infty}_{>t}|\theta_t}\Bigg[&\sum_{k=t}^{\tau(\hat{\theta}^{t-1},\theta_t,\tilde{\theta}^{\infty}_{>t})} \delta^{k-t} J_t^k(\theta_t, \tilde{\theta}^k_{>t})\psi'(\xi^*_k(\hat{\theta}^{t-1}, \theta_t, \tilde{\theta}^k_{>t})) \\ &- \sum_{k=t}^{\tau(\hat{\theta}^{t-1},\hat{\theta}_t,\tilde{\theta}^{\infty}_{>t})} \delta^{k-t} J_t^k(\theta_t, \tilde{\theta}^k_{>t})\psi'(\xi^*_k(\hat{\theta}^{t-1}, \hat{\theta}_t, \tilde{\theta}^k_{>t}))\Bigg](\theta_t - \hat{\theta}_t) \geq 0.\end{aligned} \tag{9}$$

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

$$x_t^*(\theta^t, \pi^t) = S_t(\theta^t) + \alpha_t(\theta^t)\pi_t \quad \text{all } t, \text{ all } \theta^t \in \Theta^t, \tag{10}$$

where $S_t(\theta^t)$ and $\alpha_t(\theta^t)$ are scalars that depend on the history of reported productivities such that, irrespective of the distribution Φ of the (zero-mean) transitory noise, the mechanism $\Omega = (\xi^*, x^*, \kappa^*)$ is incentive compatible and sequentially individually rational and maximizes the firm's profits. Furthermore, any contract that is incentive compatible and sequentially individually rational and maximizes the firm's profits implements the policies (ξ^*, κ^*) with probability one (i.e., except over a zero-measure set of histories).

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 and in subsequent periods, where the average is over future histories, weighted by the impulse responses. These conditions are trivially satisfied when the effort and retention policies are *strongly monotone,* that is, when each $\xi_t^*(\cdot)$ and $\kappa_t^*(\cdot)$ is nondecreasing in θ^t.[38] More generally, the conditions in the propositions require only that the expected sum of marginal disutilities of effort, conditional on retention and weighted by the impulse responses, changes sign only once when the manager changes his report about current productivity.

Turning to the components of the linear scheme, the coefficients α_t are chosen so as to provide the manager with the right incentives to choose effort obediently. Because neither future cash flows nor future retention decisions depend on current cash flows (and, as a result, on current effort), it is easy to see that, when the sensitivity of the manager's compensation to the current cash flows is given by $\alpha_t = \psi'(\xi_t^*(\theta^t))$, by choosing effort $e_t = \xi_t(\theta^t)$, the manager equates the marginal disutility of effort to its marginal benefit and hence maximizes his continuation payoff. This is irrespective of whether or not the manager has reported his productivity truthfully. Under the proposed scheme, the moral hazard part of the problem is thus controlled entirely through the variable components α_t.

Given α_t, the fixed components S_t are then chosen to control for the adverse-selection part of the problem, that is, to induce the managers to reveal their productivity. As we show in Appendix A, when the policies ξ and κ satisfy the single-crossing conditions in the proposition, then when the two components α and S are considered together, the following prop-

[38] The expression "strongly monotone" is used in the dynamic mechanism design literature to differentiate this form of monotonicity from other weaker notions (see, e.g., Courty and Li 2000; Pavan et al. 2012).

erty holds: In the continuation game that starts with any arbitrary history $h_t = (\theta^t, \hat{\theta}^{t-1}, e^{t-1}, \pi^{t-1})$, irrespective of whether or not the manager has been truthful in the past, he finds one-stage deviations from the truthful and obedient strategy unprofitable. Together with a certain property of continuity at infinity discussed in Appendix A, this result in turn implies that no other deviations are profitable either.

In a moment, we turn to primitive conditions that guarantee that the policies (ξ^*, κ^*) of proposition 3 satisfy the single-crossing conditions of proposition 4. Before doing so, we notice that, under reasonable conditions, the linear schemes of proposition 4 entail a nonnegative payment to the manager in every period and for any history. We conclude that neither our positive nor our normative results below depend critically on our simplifying assumption of disregarding limited liability (or cash) constraints.

Corollary 1. When (i) the lower bound $\underline{v}$ on the transitory noise shocks ν is not too small (i.e., not too large in absolute value), (ii) the level of the outside option U^o is not too small, and (iii) the discount factor δ is not too high, the linear schemes of proposition 4 can be chosen so as to entail a nonnegative payment to the manager in every period and for any history. Under these additional assumptions, the corresponding mechanism $\Omega = (\xi^*, x^*, \kappa^*)$ remains optimal also in settings in which the managers are protected by limited liability.

We now turn to primitive conditions that guarantee that the policies (ξ^*, κ^*) that solve the relaxed program satisfy the conditions of proposition 4 and hence are sustained under any optimal mechanism.

Proposition 5. A sufficient condition for the policies (ξ^*, κ^*) of proposition 3 to satisfy the single-crossing conditions of proposition 4 (and hence to be part of an optimal mechanism) is that, for each t, the function $\eta(\cdot)J_1^t(\cdot)$ is nonincreasing on Θ^t.[39] When this is the case, the optimal retention policy takes the form of a *cutoff rule*: There exists a sequence of nonincreasing threshold functions $(\theta_t^*(\cdot))_{t=1}^{\infty}$, $\theta_t^* : \Theta^{t-1} \to \mathbb{R}$, all $t \geq 1$, such that, conditional on being employed in period t, the manager is retained at the end of period t if and only if $\theta_t \geq \theta_t^*(\theta^{t-1})$.[40] Furthermore, under the above conditions, in each period $t \geq 1$, the optimal effort policy $\xi_t^*(\cdot)$ is nondecreasing in the reported productivities.

Note that the monotonicity condition in the proposition guarantees that each $\xi_t^*(\theta^t)$ is nondecreasing, which is used to guarantee implement-

[39] With bounded noise v, the monotonicity condition in the proposition can be replaced by the weaker condition that $\theta_t - \eta(\theta_1) J_1^t(\theta^t)$ be nondecreasing in θ^t for all t. Under this condition, the policies (ξ^*, κ^*) remain implementable (albeit not necessarily with linear schemes), and the results in the proposition continue to hold. The same is true for some, but not all, unbounded noise distributions.

[40] The cutoff θ_1^* is a scalar.

ability in linear schemes. It also guarantees that the flow virtual surplus

$$VS_t(\theta^t) \equiv \xi^*_t(\theta^t) + \theta_t - \psi(\xi^*_t(\theta^t)) - \eta(\theta_1)J^t_1(\theta^t)\psi'(\xi^*_t(\theta^t)) - (1-\delta)U^o \quad (11)$$

that the firm expects from each incumbent during the tth period of employment is nondecreasing in the history of productivities θ^t. Together with the condition of "first-order stochastic dominance in types" (which implies that impulse responses are nonnegative), this property in turn implies that the value $W^*(\theta^t)$ of continuing the relationship after t periods is nondecreasing. In this case, the turnover policy κ^* that maximizes the firm's virtual surplus is also nondecreasing and takes the form of a simple cutoff rule, with cutoff functions $(\theta^*_t(\cdot))^{\infty}_{t=1}$ satisfying the properties in the proposition.

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

Definition 1. The kernels F satisfy the property of *declining impulse responses* if, for any $t > s \geq 1$, any $(\theta^s, \theta^t_{>s})$, $\theta_t \geq \theta_s$ implies that $J^t_1(\theta^s, \theta^t_{>s}) \leq J^s_1(\theta^s)$.

As anticipated in the introduction, this property captures the idea that the effect of a manager's initial productivity on his future productivity declines with the length of the employment relationship, a property that seems reasonable for many cases of interest. This property is satisfied, for example, by an autonomous AR(1) process $\theta_t = \gamma\theta_{t-1} + \varepsilon_t$ with coefficient γ of linear dependence smaller than one.

We then have the following result.

Proposition 6. Suppose that, for each t, the function $\eta(\cdot)J^t_1(\cdot)$ is nonincreasing on Θ^t. Suppose in addition that the kernels F satisfy the property of declining impulse responses. Take an arbitrary period $t \geq 1$ and any $\theta^t \in \Theta^t$ such that $\kappa^*_{t-1}(\theta^{t-1}) = 1$. If θ^t is such that $\theta_t \geq \theta_s$ for some $s < t$, then $\kappa^*_t(\theta^t) = 1$.

In words, when separation occurs, it must necessarily be the case that the manager's productivity is at its historical lowest. Along with the result in proposition 5 that the threshold functions $\theta^*_t(\cdot)$ are nonincreasing, this result implies that the productivity level that the firm requires for retention declines with the length of the employment relationship.[41] The reason why the retention policy becomes gradually more permissive over time is the one anticipated in the introduction. Suppose that the effect

[41] That is, the threshold functions $(\theta^*_t(\cdot))^{\infty}_{t=1}$ that describe the optimal retention policy must satisfy the following property: for any $t \geq 1$, any $\theta^t = (\theta_s)^t_{s=1}$ with $\theta_s \geq \theta^*_s(\theta^{s-1})$ all $s \leq t$, necessarily $\theta^*_{t+1}(\theta^t) \leq \theta^*_t(\theta^{t-1})$. Note that this also implies that there exists a nondecreasing sequence of scalars $(\theta^{\#}_t)^{\infty}_{t=1}$ such that a manager is retained in period 1 if and only if $\theta_1 \geq \theta^{\#}_1$ and, for any $t \geq 2$, no manager whose period t productivity is above $\theta^{\#}_t$ is fired in period t; this can be seen by letting

$$\theta^{\#}_t = \max\{\theta^*_t(\theta^{t-1}) : \theta^{t-1} \text{ satisfies } \theta_s \geq \theta^*_s(\theta^{s-1}) \text{ all } s \leq t-1\}.$$

of the initial productivity on future productivity declines over time, and consider a manager whose initial type is θ_1. A commitment to replace this manager in the distant future is less effective in reducing the informational rent that the firm must leave to each more productive type $\theta_1' > \theta_1$ than a commitment to replace him in the near future (for given productivity at the time of dismissal). Formally, for any given productivity $\theta \in \Theta$, the net flow payoff that the firm expects (ex ante) from retaining the incumbent in period t, as captured by (11), increases with the length of the employment relationship, implying that the value function W^* increases as well.

Remark 1. Note that, while the result in proposition 6 is reinforced by the fact that, under the optimal contract, effort increases over time, it is not driven by this property. The same result would hold if the level of effort that the firm asks of the manager were exogenously fixed at some constant level $\hat{e}$.

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 who are highly productive at the early stages (and hence, on average, generate higher profits) by offering them job stability once their tenure in the firm becomes long enough. Thus, what in the eyes of an external observer may look like "entrenchment" can actually be the result of a profit-maximizing contract in a world in which managerial productivity is expected to change over time and to be the managers' private information. Importantly, note that this property holds independently of the level of the managers' outside option U^o. We thus expect such a property to hold irrespective of whether one looks at a given firm or at the entire market equilibrium.

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

V. On the (In)Efficiency of Profit-Maximizing Retention Decisions

We now turn to the normative implications of the result that profit-maximizing retention policies become more permissive with time. We start by establishing that the first-best effort and turnover policies of proposition 1 remain implementable also when productivity and effort choices are the managers' private information.

Proposition 7. Assume that both productivity and effort choices are the managers' private information. There exists a linear compensation scheme of the type described in proposition 4 that implements the first-best effort and turnover policies of proposition 1.

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 under the firm's optimal contract gradually converges to its efficient level as the length of the employment relationship grows sufficiently large. One might expect a similar convergence result to apply also to retention decisions. This conjecture, however, fails to take into account that the firm's endogenous separation payoff (i.e., the payoff that the firm expects from going back to the labor market and offering the profit-maximizing contract to each new manager) is lower than the planner's endogenous separation payoff (i.e., the surplus that the planner expects by forcing the firm to go back to the labor market and offer the welfare-maximizing contract to each new manager). Taking this into account, one can then show that, once the length of the employment relationship has grown sufficiently large, profit-maximizing retention decisions become excessively permissive as compared to what efficiency requires. We formalize this result in proposition 8 below. Before doing that, as a preliminary step toward understanding the result, we consider a simplified example.[42]

Example 1. Consider a firm operating for only two periods, and assume that this is commonly known. In addition, suppose that both θ_1 and ε_2 are uniformly distributed over $[-.5, +.5]$ 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 $\gamma < 0.845$), where $J_1^2 = \gamma$ is the impulse response of θ_2 to θ_1.

The relation between the profit-maximizing thresholds θ_1^* and the impulse response γ of θ_2 to θ_1 is depicted in figure 1 (the efficient threshold is $\theta^E = 0$).

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 of θ_2 to θ_1. When γ

[42] The reader may notice that this example fails to satisfy the assumption that each kernel has the same support. However, recall that such an assumption was made only to simplify the exposition. All our results extend to processes with shifting supports, as well as to nonautonomous processes (see the working paper version of this paper, Garrett and Pavan [2011*b*]).

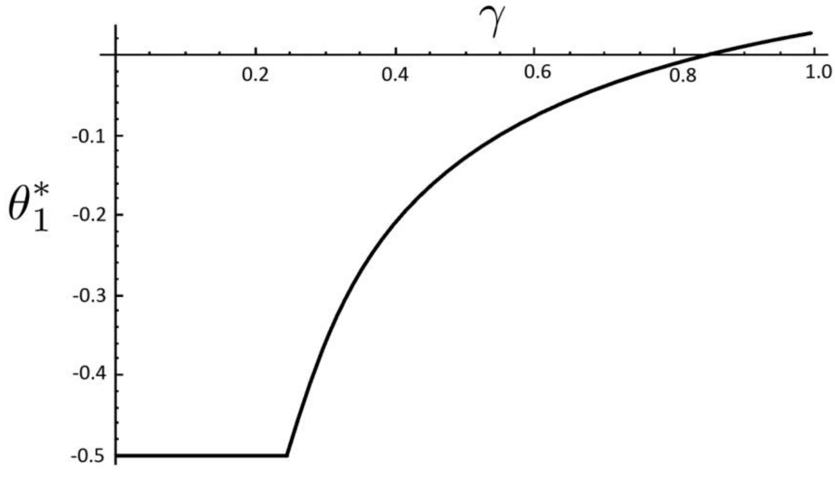

FIG. 1.—Relation between retention thresholds and impulse responses

is small, the effect of θ_1 on θ_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 1 to retaining the incumbent for a large set of his period 1 productivities. In particular, when γ is very small (i.e., when θ_1 and θ_2 are close to being independent), the firm optimally commits to retaining the incumbent irrespective of his period 1 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 $\theta_1 \geq \theta^E = 0$.

On the other hand, when γ is close to one, 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, that is, 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 virtual surplus functions VS_1 and VS_2 are given by (11). In this example, VS_1 is strictly convex. Noting that $\theta^E = \mathbb{E}[\tilde{\theta}_1]$, we then have that

$$\mathbb{E}[VS_1(\tilde{\theta}_1)] > VS_1(\theta^E) = \mathbb{E}_{\tilde{\theta}_2|\theta^E}[VS_2(\theta^E, \tilde{\theta}_2)];$$

that is, the expected value of replacing the incumbent is greater than the value of keeping him when his first-period productivity equals the efficient threshold. The same result holds for γ close to one. When productivity is highly persistent, the firm's optimal contract may thus induce excessive firing (equivalently, too high a level of turnover) as compared to what is efficient.

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

DEFINITION 2. The kernels F satisfy the property of *vanishing impulse responses* if, for any $\epsilon > 0$, there exists t_ϵ such that, for all $t > t_\epsilon$, $\eta(\theta_1)J_1^t(\theta^t) < \epsilon$ for all $\theta^t \in \Theta^t$.

This condition simply says that the effect of the managers' initial productivity on their subsequent productivity eventually vanishes after sufficiently long tenure and that this occurs uniformly over all histories.

Next, we introduce an additional technical condition that plays no substantial role but permits us to state our key normative result in the cleanest possible manner.

CONDITION LC (Lipschitz continuity). (a) There exists a constant $\beta \in \mathbb{R}_{++}$ such that, for each $t \geq 2$, each $\theta_1 \in \Theta$, the function $\eta(\theta_1)J_1^t((\theta_1, \cdot))$ is Lipschitz continuous over Θ^{t-1} with Lipschitz constant β; and (b) there exists a constant $\rho \in \mathbb{R}_{++}$ such that, for $\theta \in \Theta$, the function $f(\theta|\cdot)$ is Lipschitz continuous over Θ with constant ρ.

We then have the following result (the result in this proposition and the result in corollary 2 below refer to the interesting case in which $\theta^E \in \text{int}\{\Theta\}$).

PROPOSITION 8. (i) Suppose that, for each t, the function $\eta(\cdot)J_1^t(\cdot)$ is nonincreasing on Θ^t. Suppose also that the kernels F satisfy the property of vanishing impulse responses. There exists $\bar{t} \in \mathbb{N}$ such that, for any $t > \bar{t}$ and any $\theta^t \in \Theta^t$ for which $\theta_t \geq \theta^E$, $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] > \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$. (ii) Suppose, in addition, that F satisfies the properties of condition LC. Then there exists $\bar{t} \in \mathbb{N}$ such that, for any $t > \bar{t}$, any $\theta^{t-1} \in \Theta^{t-1}$ for which $\kappa_{t-1}^*(\theta^{t-1}) = 1$, $\theta_t^*(\theta^{t-1}) < \theta^E$.

Part i of proposition 8 establishes existence of a critical length $\bar{t}$ for the employment relationship after which retention is excessive under the profit-maximizing contract. For any $t > \bar{t}$, any $\theta^t \in \Theta^t$, if the manager is retained at the end of period t under the efficient contract, he is also retained under the profit-maximizing contract. Condition LC implies continuity in θ_t of the expected continuation payoffs $\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{t-1},\cdot)}[W^*(\tilde{\theta}^{t+1})]$ and $\mathbb{E}_{\tilde{\theta}|\cdot}[W^E(\tilde{\theta})]$ for any period $t \geq 2$ and history of productivities $\theta^{t-1} \in \Theta^{t-1}$. This in turn establishes that the profit-maximizing retention thresholds will eventually become strictly smaller than their efficient counterparts (as stated by part ii).

The proof for proposition 8 can be understood heuristically by considering the "fictitious problem" that involves maximizing the firm's expected profits in a setting in which the firm can observe its incumbent manager's types and effort choices but not those of its future hires. In this environment, the firm optimally asks the incumbent to follow the efficient effort policy in each period, it extracts all surplus from the incumbent (i.e., the incumbent receives a payoff equal to his outside option), and it offers the contract identified in proposition 3 to each new hire.

Now, consider the actual problem. After a sufficiently long tenure, the cutoffs for retaining the incumbent in this problem must converge to those in the fictitious problem. The reason is that, after a sufficiently long tenure, distorting effort and retention decisions has almost no effect on the ex ante surplus that the firm must leave to the incumbent. 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 in which 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, whereas this is possible with symmetric information. It follows that, after a sufficiently long tenure, the value that 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 that the firm obtains under the incumbent's control is the same, whereas the payoff from hiring a new manager is lower. Furthermore, because the value that 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,[43] we have that the firm's retention policy necessarily becomes more permissive than the efficient one after sufficiently long tenure.

The findings of propositions 6 and 8 can be combined together to establish the following corollary, which contains our key normative result. (The result refers to the interesting case in which the profit-maximizing policy retains each manager after the first period with positive probability, that is, $\theta_1^* < \bar{\theta}$.)

Corollary 2. Suppose that, in addition to satisfying the property that, for each t, the function $\eta(\cdot)J_1^t(\cdot)$ is nonincreasing on Θ^t, the kernels F satisfy the properties of both declining and vanishing impulse responses. Then, relative to what is efficient, the profit-maximizing contract induces either excessive retention (i.e., too little turnover) throughout the entire relationship or excessive firing at the early stages followed

[43] Recall that welfare under the efficient contract with asymmetric information coincides with the sum of the firm's expected profits and of all the managers' outside options under the contract that the firm would offer if information about all managers' effort and productivities were symmetric.

by excessive retention in the long run. Formally, there exist dates $\underline{t}, \bar{t} \in \mathbb{N}$, with $1 \leq \underline{t} \leq \bar{t}$, such that ($a$) for any $t < \underline{t}$ and almost any $\theta^t \in \Theta^t$, if $\kappa^E_{t-1}(\theta^{t-1}) = 1$ and $\kappa^E_t(\theta^t) = 0$, then $\kappa^*_t(\theta^t) = 0$; and (b) for any $t > \bar{t}$ and almost any $\theta^t \in \Theta^t$, if $\kappa^*_{t-1}(\theta^{t-1}) = 1$ and $\kappa^*_t(\theta^t) = 0$, then $\kappa^E_t(\theta^t) = 0$.

Hence, any manager who is fired at the end of period $t < \underline{t}$ under the efficient policy is fired either 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 fired either at the end of the same period or earlier under the efficient policy.

VI. Conclusions

We developed a tractable, yet rich, model of dynamic managerial contracting that explicitly accounts for the following possibilities: (i) turnover is driven by variations in the managers' ability to generate profits for the firm (equivalently, in the match quality); (ii) variations in managerial productivity are anticipated at the time of contracting but privately observed by the managers; (iii) at each point in time, the firm can go back to the labor market and replace an incumbent manager with a new hire; and (iv) the firm's prospects under the new hire are affected by the same information frictions as in the relationship with each incumbent.

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

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

Appendix A

Proof of Proposition 1

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

Consider the retention policy. Because all managers are ex ante identical and because the process governing the evolution of the managers' productivities is Markov and autonomous, it is immediate that, in each period, the decision of whether or not to retain a manager must depend only on the manager's current productivity θ. We will denote by $W^E : \Theta \to \mathbb{R}$ the value function associated with the problem that involves choosing the efficient Markovian retention policy, given the constant effort policy described above. For any $\theta \in \Theta$, $W^E(\theta)$ specifies the maximal continuation expected welfare that can be achieved when the incumbent manager's productivity is θ. 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 defined in the main text.

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

Note that the set $\mathcal{N}^E$, together with the uniform metric, is a complete metric space. Because the process satisfies the property of first-order-stochastic dominance in past types, $\mathcal{N}^E$ is closed under T_E. Moreover, "Blackwell's sufficient conditions" (namely, "monotonicity" and "discounting," where the latter is guaranteed by the assumption that $\delta < 1$) imply that T_E is a contraction. Therefore, by the contraction mapping theorem (see, e.g., theorem 3.2 of Stokey and Lucas [1989]), for any $W \in \mathcal{N}^E$, $\hat{W}^E = \lim_{n\to\infty} T_E^n W$ exists, is unique, and belongs to $\mathcal{N}^E$.

Now, we claim that the following retention policy is efficient: for any t, any $\theta^t \in \Theta^t$, $\kappa_{t-1}(\theta^{t-1}) = 1$ implies $\kappa_t(\theta^t) = 1$ if $\mathbb{E}_{\tilde{\theta}|\theta_t}[\hat{W}^E(\tilde{\theta})] \geq \mathbb{E}_{\tilde{\theta}_1}[\hat{W}^E(\tilde{\theta}_1)]$ and $\kappa_t(\theta^t) = 0$ otherwise. Note that, because the process satisfies the property of first-order stochastic dominance in past types and because $\hat{W}$ is nondecreasing, this retention policy is a cutoff policy. This property, together with the fact that the "flow payoffs" $\theta + e^E - \psi(e^E) - (1-\delta)U^o$ and $\hat{W}^E$ are uniformly bounded on Θ, then permit one to verify, via standard verification arguments, that the constructed policy is indeed efficient and that $W^E = \hat{W}^E$.[44] QED

Proof of Proposition 3

First, consider the effort policy. It is easy to see that the policy ξ^* that solves the relaxed program is independent of the retention policy κ and is such that $\xi_t^*(\theta^t)$ is given by (8) for all t, all $\theta^t \in \Theta^t$. Next, consider the retention policy. We first prove existence of a unique fixed point $W^* \in \mathcal{B}$ to the mapping $T(\xi^*)$. To this end, endow $\mathcal{B}$ with the uniform metric. That $\mathcal{B}$ is closed under $T(\xi^*)$ is ensured by the restrictions on ψ and by the definition of ξ^*, which together imply that each function $VS_t : \Theta^t \to \mathbb{R}$ defined by

[44] This verification is standard in dynamic programming and hence is omitted for brevity.

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

is uniformly bounded over A. Blackwell's theorem implies that $T(\xi^*)$ is a contraction mapping, and the contraction mapping theorem (see Stokey and Lucas 1989) then implies the result. Standard arguments then permit one to verify that $W^*(\theta^t)$ is indeed the value function associated with the problem that involves choosing a retention policy that, given the history of productivities $\theta^t \in \Theta^t$ for the incumbent manager and given the profit-maximizing effort policy ξ^*, maximizes the firm's expected total continuation profits.[45] Having established this result, it is then easy to see that any retention policy κ^* that, given the effort policy ξ^*, maximizes the firm's total profits must satisfy the conditions in the proposition. QED

Proof of Proposition 4

Consider the linear reward scheme $x = (x_t : \Theta^t \times \mathbb{R} \to \mathbb{R})_{t=1}^{\infty}$, where $x_t(\theta^t, \pi_t) = S_t(\theta^t) + \alpha_t(\theta^t)\pi_t$ for all t, with

$$\alpha_t(\theta^t) \equiv \psi'(\xi_t^*(\theta^t)) \tag{A1}$$

and

$$\begin{aligned} S_t(\theta^t) = {} & \psi(\xi_t^*(\theta^t)) - \alpha_t(\theta^t)[\xi_t^*(\theta^t) + \theta_t] + (1-\delta)U^o \\ & + \int_{\underline{\theta}}^{\theta_t} \mathbb{E}_{\tilde{\theta}^{\infty}_{>t}|s}\left[\sum_{k=t}^{\tau(\theta^{t-1},s,\tilde{\theta}^{\infty}_{>t})} \delta^{k-t} J_t^k(s, \tilde{\theta}^k_{>t})\psi'(\xi_k^*(\theta^{t-1}, s, \tilde{\theta}^k_{>t}))\right] ds \\ & - \delta\kappa_t^*(\theta^t)\mathbb{E}_{\tilde{\theta}_{t+1}|\theta_t}[u_{t+1}(\tilde{\theta}_{t+1}; \theta^t)], \end{aligned} \tag{A2}$$

where

$$\begin{aligned} u_{t+1}(\theta_{t+1}; \theta^t) \equiv {} & \int_{\underline{\theta}}^{\theta_{t+1}} \mathbb{E}_{\tilde{\theta}^{\infty}_{>t+1}|s}\left[\sum_{k=t+1}^{\tau(\theta^t,s,\tilde{\theta}^{\infty}_{>t+1})} \delta^{k-(t+1)} J_{t+1}^k(s, \tilde{\theta}^k_{>t+1}) \right. \\ & \left. \times \ \psi'(\xi_k^*(\theta^t, s, \tilde{\theta}^k_{>t+1}))\right] ds \end{aligned} \tag{A3}$$

denotes the manager's period $t+1$ continuation payoff (over and above his outside option) under the truthful and obedient strategy.

Note that, because retention does not depend on cash flows, it does not affect the manager's incentives for effort. From the law of iterated expectations, it then follows that, for any given history of reports $\hat{\theta}^{t-1}$ such that the manager is still employed in period $t \geq 1$ (i.e., $\kappa_{t-1}^*(\hat{\theta}^{t-1}) = 1$) and for any period t productivity θ_t, the manager's continuation payoff at the beginning of period t when the manager plans to follow a truthful and obedient strategy from period t onward is given by $U^o + u_t(\theta_t; \hat{\theta}^{t-1})$, where[46]

[45] The reason why the term $-(1-\delta)U^o$ disappears from the mapping $T(\xi^*)$ is that this term is constant across t and across all managers.

[46] Note that, under the proposed scheme, a manager's continuation payoff depends on past announcements $\hat{\theta}^{t-1}$, but not on past productivities θ^{t-1}, effort choices e^{t-1}, or cash flows π^{t-1}.

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

Because $u_t(\theta_t;\hat{\theta}^{t-1})\geq 0$, the above scheme guarantees that, after any truthful and obedient history, the manager finds it optimal to stay in the relationship whenever the firm's retention policy permits him to do so.

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

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

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 onward gives type θ_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 $t+1$ onward.

Now, to establish the result in the proposition, it suffices to compare the manager's continuation payoff at any period t, given any possible type θ_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 onward, with the manager's expected payoff under any continuation strategy that satisfies the following property. In each period $s\geq t$ and after any possible history of reports $\hat{\theta}^s\in\Theta^s$, the effort specified by the strategy for period s coincides with the one prescribed by the recommendation policy ξ_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

$$\psi'(\xi_s(\hat{\theta}^s))=\alpha_s(\hat{\theta}^s). \tag{A4}$$

Restricting attention to continuation strategies in which, at any period $s\geq t$, the manager follows the recommended effort policy $\xi_s(\hat{\theta}^s)$ is justified by (i) the fact that the compensation paid in each period $s\geq t$ is independent of past cash flows π^{s-1}; (ii) under the proposed scheme, the manager's period s compensation, net of his disutility of effort, is maximized at $e_s=\xi_s(\hat{\theta}^s)$; and (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 (A4), there exists another continuation strategy whose effort choices comply with (A4) for all $s \geq t$, all $\hat{\theta}^s$, which gives the manager a (weakly) higher expected continuation payoff.

Next, it is easy to see that, under any continuation strategy that satisfies the aforementioned effort property, the manager's expected payoff in each period $s \geq t$ is bounded uniformly over Θ^s. In turn, this implies that a continuity-at-infinity condition similar to that in Fudenberg and Levine (1983) holds in this environment. Precisely, for any $\epsilon > 0$, there exists t large enough such that, for all $\theta_t \in \Theta_t$ and all $\hat{\theta}^{t-1}, \breve{\theta}^{t-1} \in \Theta^{t-1}$, $\delta^t|\hat{u}_t(\theta_t; \hat{\theta}^{t-1}) - \breve{u}_t(\theta_t; \breve{\theta}^{t-1})| < \epsilon$, where $\hat{u}_t$ and $\breve{u}_t$ are continuation payoffs under arbitrary continuation strategies satisfying the above effort restriction, given arbitrary histories of reports $\hat{\theta}^{t-1}$ and $\breve{\theta}^{t-1}$. This continuity-at-infinity property, together with the aforementioned property about one-stage deviations from a truthful and obedient strategy, imply that, after any history, the manager's continuation payoff under a truthful and obedient strategy from that period onward is weakly higher than the expected payoff under any other continuation strategy. We thus conclude that, whenever the pair of policies (ξ^*, κ^*) satisfies all the single-crossing conditions in the proposition, it can be implemented by the proposed linear reward scheme. That is, the mechanism $\Omega^* = (\xi^*, \kappa^*, s^*)$ is incentive compatible and sequentially individually rational.

That the mechanism Ω^* is optimal then follows from proposition 2 by observing that, under Ω^*, type $\underline{\theta}$ obtains an expected payoff equal to his outside option, that is, $V^{\Omega^*}(\underline{\theta}) = U^o$. The last claim in the proposition that the policies (ξ^*, κ^*) are implemented under any mechanism that is optimal for the firm then follows from the fact that such policies are the "essentially" unique policies that solve the relaxed program, where essentially means up to a zero-measure set of histories. QED

Proof of Corollary 1

The result follows from inspecting the terms S_t and α_t of the linear scheme defined in the proof of proposition 4. QED

Proof of Proposition 5

Assume that each function $h_t(\cdot) \equiv -\eta(\cdot)J_1^t(\cdot)$ is nondecreasing. Because the function

$$g(e, h, \theta) \equiv e + \theta - \psi(e) + h\psi'(e) - (1 - \delta)U^o$$

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 t, each θ^t, $\xi_t^*(\theta^t) = \arg\max_{e \in E} g(e, h_t(\theta^t), \theta_t)$.

Next, we show that, for all t, the function $W^*(\theta^t)$ is nondecreasing. To this aim, let $\mathcal{N} \subset \mathcal{B}$ denote the set of all bounded functions from $A \equiv \cup_{t=1}^{\infty} \Theta^t$ to $\mathbb{R}$ that, for each t, are nondecreasing in θ^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 managers' productivities satisfies the property of first-order stochastic dominance in past types, implies that $\mathcal{N}$ is closed under the operator $T(\xi^*)$. It follows that $\lim_{n\to\infty} T(\xi^*)^n W$ is in $\mathcal{N}$. The fact that $T(\xi^*) : \mathcal{B} \to \mathcal{B}$ admits a unique fixed point then implies that $\lim_{n\to\infty} T(\xi^*)^n W = W^*$.

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

$$\theta_t^*(\theta^{t-1}) = \underline{\theta} \quad \text{if } \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] > \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)] \text{ for all } \theta_t \in \Theta,$$

$$\theta_t^*(\theta^{t-1}) = \bar{\theta} \quad \text{if } \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] < \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)] \text{ for all } \theta_t \in \Theta,$$

$$\theta_t^*(\theta^{t-1}) = \min\{\theta_t \in \Theta : \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] \geq \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]\}$$
$$\text{if } \{\theta_t \in \Theta : \mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]\} \neq \varnothing .$$

The property that each $\xi_t^*(\cdot)$ and $\kappa_t^*(\cdot)$ is nondecreasing implies that the policies ξ^* and κ^* satisfy all the single-crossing conditions of proposition 4. QED

Proof of Proposition 6

We prove the proposition by showing that, for any arbitrary pair of periods s, t with $s < t$, and an arbitrary history of productivities $\theta^t = (\theta^s, \theta^t_{>s}) \in \Theta^t$, $\theta_s \leq \theta_t$ implies that $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] \geq \mathbb{E}_{\tilde{\theta}^{s+1}|\theta^s}[W^*(\tilde{\theta}^{s+1})]$.

Let $\hat{\mathcal{N}}$ denote the subclass of all functions $W \in \mathcal{B}$ satisfying the following properties: (a) for each s, $W(\theta^s)$ is nondecreasing over Θ^s; and (b) for any $t > s$, any $\theta^s \in \Theta^s$, and any θ^t such that $\theta^t = (\theta^s, \theta^t_{>s}) \in \Theta^t$, if $\theta_s \leq \theta_t$, then $W(\theta^s) \leq W(\theta^t)$.

We established already in the proof of proposition 5 that the operator $T(\xi^*)$ preserves property a. The property of declining impulse responses, together with the property of first-order stochastic dominance in past types, implies that $T(\xi^*)$ also preserves property b. The unique fixed point W^* to the mapping $T(\xi^*) : \mathcal{B} \to \mathcal{B}$ thus satisfies properties a and b above. First-order stochastic dominance in past types then implies that $\mathbb{E}_{\tilde{\theta}^{t+1}|\theta^t}[W^*(\tilde{\theta}^{t+1})] \geq \mathbb{E}_{\tilde{\theta}^{s+1}|\theta^s}[W^*(\tilde{\theta}^{s+1})]$. QED

Proof of Proposition 7

The result follows from the same arguments as in the proof of proposition 4 by observing that the first-best policies are nondecreasing. QED

Proof of Example 1

Note that $\eta(\theta_1) = \frac{1}{2} - \theta_1$. Thus, $\xi_1^*(\theta_1) = \frac{1}{2} + \theta_1$ and the payoff from hiring a new manager in period 2 is

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

The manager is thus retained if and only if

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

where $\xi_2^*(\theta_1) = 1 - (\gamma/2) + \gamma\theta_1$ and $\mathbb{E}_{\tilde{\theta}_2|\theta_1}[\tilde{\theta}_2] = \gamma\theta_1$. The inequality holds for all $\theta_1 \in \left[-\frac{1}{2}, +\frac{1}{2}\right]$ if $\gamma \leq 0.242$. Otherwise it holds if and only if $\theta_1 \geq \theta_1^*$ for some $\theta_1^* \in \left(-\frac{1}{2}, +\frac{1}{2}\right)$ such that $\theta_1^* < 0$ if $\gamma \in (0.242, 0.845)$ and $\theta_1^* > 0$ if $\gamma > 0.845$. QED

Proof of Proposition 8

The proof follows from five lemmas. Lemmas A1–A3 establish part i of the proposition. Lemmas A4 and A5, together with part i, establish part ii.

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

LEMMA A1. For any $c \in \mathbb{R}$, there exists a unique function $W^{E,c} \in \mathcal{B}^E$ that is a fixed point to the mapping $T_{E,c} : \mathcal{B}^E \to \mathcal{B}^E$ defined, for all $W \in \mathcal{B}^E$, all $\theta \in \Theta$, by

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

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

$$\mathbb{E}_{\tilde{\theta}|\theta}[W^{E,c''}(\tilde{\theta})] \geq c'' \Rightarrow \mathbb{E}_{\tilde{\theta}|\theta}[W^{E,c'}(\tilde{\theta})] > c' + \iota.$$

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 θ to $\mathbb{R}$ such that, for all $\theta \in \Theta$, $W(\theta) \geq W^{E,c''}(\theta) - \delta(c'' - c')$. 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 $\theta \in \Theta$,

$$\begin{aligned} T_{E,c'}W(\theta) - W^{E,c''}(\theta) &= T_{E,c'}W(\theta) - T_{E,c''}W^{E,c''}(\theta) \\ &= \delta(\max\{\mathbb{E}_{\tilde{\theta}|\theta}[W(\tilde{\theta})];\ c'\} - \max\{\mathbb{E}_{\tilde{\theta}|\theta}[W^{E,c''}(\tilde{\theta})];\ c''\}) \\ &\geq -\delta(c'' - c'). \end{aligned}$$

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 $\theta \in \Theta$, $W^{E,c'}(\theta) - W^{E,c''}(\theta) \geq -\delta(c'' - c')$.

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

$$\begin{aligned} \mathbb{E}_{\tilde{\theta}|\theta}[W^{E,c'}(\tilde{\theta})] &\geq \mathbb{E}_{\tilde{\theta}|\theta}[W^{E,c''}(\tilde{\theta})] - \delta(c'' - c') \\ &\geq c'' - \delta(c'' - c') > c' + \iota \end{aligned}$$

for some $\iota > 0$. QED

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

LEMMA A2. $\mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1)] > \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$.

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

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

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

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

$$\begin{aligned} W^*(\theta_1) &= T(\xi^*)W^*(\theta_1) \\ &= \xi_1^*(\theta_1) + \theta_1 - \psi(\xi_1^*(\theta_1)) - \eta(\theta_1)\psi'(\xi_1^*(\theta_1)) \\ &\quad - (1-\delta)U^o + \delta\max\{\mathbb{E}_{\tilde{\theta}^2|\theta_1}[W^*(\tilde{\theta}^2)];\ \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]\} \\ &< \theta_1 + e^E - \psi(e^E) - (1-\delta)U^o \\ &\quad + \delta\max\{\mathbb{E}_{\tilde{\theta}|\theta_1}[W^E(\tilde{\theta})];\ \mathbb{E}_{\tilde{\theta}_1}[W^E(\tilde{\theta}_1)]\} \\ &= W^E(\theta_1), \end{aligned}$$

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

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 $\theta^t \in \Theta^t$,

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

Proof of lemma A3. Recall that $W^{E,c'}$, as defined in lemma A1, is the value function for the stopping problem with efficient flow payoffs $\theta_t + e^E - \psi(e^E) - (1-\delta)U^o$ and exogenous separation payoff c'. Now let $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$. Below, we will compare the function $W^{E,c'}$ with the value function W^* associated with the profit-maximizing stopping problem. Recall that the latter is a stopping problem with flow payoffs, for each t and each θ^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 separation payoff $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$. By the property of "vanishing impulse responses," for any $\omega > 0$, there exists $\bar{t}$ such that, for any $t > \bar{t}$, any $\theta^t \in \Theta^t$,

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

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 ω than the corresponding flow payoff in the stopping problem with efficient flow payoffs and exogenous separation payoff $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$. In terms of value functions, this implies that, for all $t > \bar{t}$, all $\theta^t \in \Theta^t$,

$$W^*(\theta^t) \geq W^{E,c'}(\theta_t) - \frac{\omega}{1-\delta}. \tag{A5}$$

To see this, consider the set $\mathcal{W} \subset \mathcal{B}$ of all bounded functions W from $A \equiv \cup_{t=1}^{\infty} \Theta^t$ to $\mathbb{R}$ such that, for all $t > \bar{t}$, all $\theta^t \in \Theta^t$, $W(\theta^t) \geq W^{E,c'}(\theta_t) - [\omega/(1-\delta)]$ and consider the operator $T_{c'} : \mathcal{B} \to \mathcal{B}$ defined, for all $t > \bar{t}$, all $\theta^t \in \Theta^t$, by

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

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

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

Since $\mathcal{W}$, together with the uniform metric, is a complete metric space and since $T_{c'}$ is a contraction, given any $W \in \mathcal{W}$, $\lim_{n\to\infty} T_{c'}^n W$ exists and belongs to $\mathcal{W}$. Furthermore, because $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$, it must be that $W^* = \lim_{n\to\infty} T_{c'}^n W$. Hence, $W^* \in \mathcal{W}$, which proves (A5).

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

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

The first inequality follows from (A5) and the second inequality follows from lemma A1 using $c' = \mathbb{E}_{\tilde{\theta}_1}[W^*(\tilde{\theta}_1)]$ and choosing ι as in that lemma. The result then follows by choosing ω sufficiently small that $\iota - [\omega/(1-\delta)] > 0$. QED

Part ii: The proof follows from two lemmas. Lemma A4 establishes Lipschitz continuity in θ_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 F satisfies the properties of condition LC. Then, for each $t \geq 2$ and each $\theta^{t-1} \in \Theta^{t-1}$, $\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{t-1},\cdot)}[W^*(\tilde{\theta}^{t+1})]$ is Lipschitz continuous over Θ. Moreover, $\mathbb{E}_{\tilde{\theta}|\cdot}[W^E(\tilde{\theta})]$ is Lipschitz continuous over Θ.

Proof of lemma A4. We show that, for any $t \geq 2$, any $\theta^{t-1} \in \Theta^{t-1}$, $\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{t-1},\cdot)}[W^*(\tilde{\theta}^{t+1})]$ is Lipschitz continuous over Θ. The proof that $\mathbb{E}_{\tilde{\theta}|\cdot}[W^E(\tilde{\theta})]$ is Lipschitz continuous over Θ is similar and is omitted. Let

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

and

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

where $K = \max\{|\underline{\theta}|, \bar{\theta}\}$ and $L > 0$ is a uniform bound on ψ'.

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

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

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

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

$$\begin{aligned} &T(\xi^*)W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) - T(\xi^*)W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}) \\ &\quad = [\xi^*_t(\theta^t) + \theta_t - \psi(\xi^*_t(\theta^t)) - \eta(\theta_1)J^t_1(\theta^t)\psi'(\xi^*_t(\theta^t)) \\ &\qquad - (1-\delta)U^o]_{\theta^t = (\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})} \\ &\qquad - [\xi^*_t(\theta^t) + \theta_t - \psi(\xi^*_t(\theta^t)) - \eta(\theta_1)J^t_1(\theta^t)\psi'(\xi^*_t(\theta^t)) \\ &\qquad - (1-\delta)U^o]_{\theta^t = (\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})} \\ &\qquad + \delta(\max\{\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})}[W(\tilde{\theta}^{t+1})]; \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1)]\} \\ &\qquad - \max\{\mathbb{E}_{\tilde{\theta}^{t+1}|(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})}[W(\tilde{\theta}^{t+1})]; \mathbb{E}_{\tilde{\theta}_1}[W(\tilde{\theta}_1)]\}). \end{aligned} \tag{A6}$$

The first two terms on the right-hand side of (A6) are no greater than $(1 + \beta L)(\theta'_\tau - \theta''_\tau)$. This can be derived as follows. For any $2 \le \tau \le t$, any $\theta^t \in \Theta^t$, any $e \in E$, define

$$g_t(\theta^t, e) = e + \theta_t - \psi(e) - \eta(\theta_1) J_1^t(\theta^t)\psi'(e) - (1-\delta)U^o.$$

For any $\theta^t = (\theta^{\tau-1}, \theta_\tau, \theta^t_{>\tau}) \in \Theta^t$, g_t is Lipschitz continuous in θ_τ and

$$\frac{\partial}{\partial \theta_\tau} g_t(\theta^{\tau-1}, \theta_\tau, \theta^t_{>\tau}, e) \le 1 + \beta L$$

for all $e \in E$ and almost all $\theta_\tau \in \Theta$. 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 of (A6) is no greater than $\delta(2\rho MK + m)(\theta'_\tau - \theta''_\tau)$. This follows because

$$\begin{aligned}
&\mathbb{E}_{\tilde\theta^{t+1}|(\theta^{\tau-1},\theta'_\tau,\theta^t_{>\tau})}[W(\tilde\theta^{t+1})] - \mathbb{E}_{\tilde\theta^{t+1}|(\theta^{\tau-1},\theta''_\tau,\theta^t_{>\tau})}[W(\tilde\theta^{t+1})] \\
&\quad = \mathbb{E}_{\tilde\theta_{t+1}|(\theta^{\tau-1},\theta'_\tau,\theta^t_{>\tau})}[W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1})] \\
&\qquad - \mathbb{E}_{\tilde\theta_{t+1}|(\theta^{\tau-1},\theta''_\tau,\theta^t_{>\tau})}[W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1})] \\
&\qquad + \mathbb{E}_{\tilde\theta_{t+1}|(\theta^{\tau-1},\theta''_\tau,\theta^t_{>\tau})}[W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1}) \\
&\qquad - W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1})] \\
&\quad = \int_\Theta W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \theta_{t+1})[f(\theta_{t+1}|\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) \\
&\qquad - f(\theta_{t+1}|\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})]\,d\theta_{t+1} \\
&\qquad + \mathbb{E}_{\tilde\theta_{t+1}|(\theta^{\tau-1},\theta''_\tau,\theta^t_{>\tau})}[W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1}) \\
&\qquad - W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}, \tilde\theta_{t+1})] \\
&\quad \le (2\rho MK + m)(\theta'_\tau - \theta''_\tau),
\end{aligned} \tag{A7}$$

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

$$\begin{aligned}
T(\xi^*)W(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau}) - T(\xi^*)W(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau}) \\
\le (1 + \beta L + 2\delta\rho MK + \delta m)(\theta'_\tau - \theta''_\tau) \\
= m(\theta'_\tau - \theta''_\tau).
\end{aligned}$$

Since $(\theta^{\tau-1}, \theta'_\tau, \theta^t_{>\tau})$ and $(\theta^{\tau-1}, \theta''_\tau, \theta^t_{>\tau})$ were arbitrary, it follows that for any $\theta_1 \in \Theta$ and any t, the function $T(\xi^*)W(\theta_1, \cdot)$ is Lipschitz continuous over $\Theta^t_{>1}$ with constant m; that is, $T(\xi^*)W$ indeed satisfies property ii above. Finally, that $T(\xi^*)W$ satisfies property iii follows from the fact that the mapping $T(\xi^*)$ preserves the monotonicity of W. We thus conclude that $T(\xi^*)W \in \mathcal{L}(M, m)$, which verifies that $\mathcal{L}(M, m)$ is closed under the $T(\xi^*)$ 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 $T(\xi^*)$ is a contraction, then implies that $W^* \in \mathcal{L}(M, m)$. Using the same argument as in (A7), we then have that $\mathbb{E}_{\tilde\theta^{t+1}|(\theta^{t-1},\cdot)}[W^*(\tilde\theta^{t+1})]$ is Lipschitz continuous over Θ with constant $(2\rho MK + m)$. QED

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 recall that we are assuming that $\theta^E \in \operatorname{int}\{\Theta\}$. The continuity of $\mathbb{E}_{\tilde\theta|\cdot}[W^E(\tilde\theta)]$ established in the previous lemma implies $\mathbb{E}_{\tilde\theta|\theta^E}[W^E(\tilde\theta)] = \mathbb{E}_{\tilde\theta_1}[W^E(\tilde\theta_1)]$. Since $t > \bar t$, by lemma A3, it follows that

$$\mathbb{E}_{\tilde\theta^{t+1}|(\theta^{t-1},\theta^E)}[W^*(\tilde\theta^{t+1})] > \mathbb{E}_{\tilde\theta_1}[W^*(\tilde\theta_1)].$$

By lemma A4, $\mathbb{E}_{\tilde\theta^{t+1}|(\theta^{t-1},\cdot)}[W^*(\tilde\theta^{t+1})]$ is continuous. Since $\theta^E \in \operatorname{int}\{\Theta\}$, there exists $\epsilon > 0$ such that, for all $(\theta^{t-1}, \theta_t) \in \Theta^t$ with $\theta_t \in (\theta^E - \epsilon, \theta^E)$, $\mathbb{E}_{\tilde\theta^{t+1}|(\theta^{t-1},\theta_t)}[W^*(\tilde\theta^{t+1})] > \mathbb{E}_{\tilde\theta_1}[W^*(\tilde\theta_1)]$. It follows that $\theta^*_t(\theta^{t-1}) < \theta^E$. QED

This concludes the proof of proposition 8.

Proof of Corollary 2

First, consider the case in which, for all $\theta_1 > \theta^E$, $\mathbb{E}_{\tilde\theta_2|\theta_1}[W^*(\tilde\theta^2)] > \mathbb{E}_{\tilde\theta_1}[W^*(\tilde\theta_1)]$. Proposition 6, together with the monotonicity property of W^* established in proposition 5, then implies that, for any $t \ge 1$, any $\theta^t \in \Theta^t$ such that $\theta_1, \theta_t > \theta^E$, $\mathbb{E}_{\tilde\theta^{t+1}|\theta^t}[W^*(\tilde\theta^{t+1})] > \mathbb{E}_{\tilde\theta_1}[W^*(\tilde\theta_1)]$. This means that, for any t, any $\theta^t \in \Theta^t$ such that $\kappa^*_{t-1}(\theta^{t-1}) = 1$ and $\kappa^*_t(\theta^t) = 0$, necessarily $\kappa^E_t(\theta^t) = 0$ (except for the possibility that θ^t is such that $\theta_s = \theta^E$ for some $s \le t$, which, however, has zero measure). That is, any manager who is fired in period t under the firm's profit-maximizing contract is fired either in the same period or earlier under the efficient contract. The result in the proposition then holds for $\underline{t} = \bar t = 1$.

Next, assume that there exists a $\theta_1 > \theta^E$ such that $\mathbb{E}_{\tilde\theta_2|\theta_1}[W^*(\tilde\theta^2)] < \mathbb{E}_{\tilde\theta_1}[W^*(\tilde\theta_1)]$, which implies that $\theta^*_1 > \theta^E$. By assumption, the manager is retained with positive probability after the first period, that is, $\theta^*_1 \in (\theta^E, \bar\theta)$. The result then holds by letting $\underline{t} = 2$. In this case, the existence of a $\bar t \ge \underline{t}$ satisfying the property in the corollary follows directly from proposition 8. QED

References

Acharya, Sankarshan. 1992. "Maximizing the Market Value of a Firm to Choose Dynamic Policies for Managerial Hiring." *Internat. Econ. Rev*. 33:373–97.

Allgood, Sam, and Kathleen Farrell. 2003. "The Match between CEO and Firm." *J. Bus*. 76:317–41.
Almazan, Andres, and Javier Suarez. 2003. "Entrenchment and Severance Pay in Optimal Governance Structures." *J. Finance* 58:519–48.
Athey, Susan, and Ilya Segal. 2007. "An Efficient Dynamic Mechanism." Manuscript, Stanford Univ.
Atkeson, Andrew, and Harold Cole. 2005. "A Dynamic Theory of Optimal Capital Structure and Executive Compensation." Manuscript, Univ. California, Los Angeles.
Banks, Jeffrey, and Rangarajan Sundaram. 1998. "Optimal Retention in Agency Problems." *J. Econ. Theory* 82:293–323.
Baron, David, and David Besanko. 1984. "Regulation and Information in a Continuing Relationship." *Information Econ. and Policy* 1:447–70.
Battaglini, Marco. 2005. "Long-Term Contracting with Markovian Consumers." *A.E.R.* 95 (3): 637–58.
Bebchuk, Lucian, and Jessie Fried. 2004. *Pay without Performance*. Cambridge, MA: Harvard Univ. Press.
Bergemann, Dirk, and Maher Said. 2011. "Dynamic Auctions." In *Wiley Encyclopedia of Operations Research and Management Science*. Hoboken, NJ: Wiley.
Bergemann, Dirk, and Juuso Välimäki. 2010. "The Dynamic Pivot Mechanism." *Econometrica* 78 (2): 771–89.
Besanko, David. 1985. "Multiperiod Contracts between Principal and Agent with Adverse Selection." *Econ. Letters* 17:33–37.
Biais, Bruno, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet. 2007. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Rev. Econ. Studies* 74 (2): 345–90.
Board, Simon. 2008. "Durable-Goods Monopoly with Varying Demand." *Rev. Econ. Studies* 75 (2): 391–413.
———. 2011. "Relational Contracts and the Value of Loyalty." *A.E.R.* 101 (7): 3349–67.
Board, Simon, and Andy Skrzypacz. 2010. "Revenue Management with Forward-Looking Buyers." Manuscript, Univ. California, Los Angeles, and Stanford Univ.
Clementi, Gian Luca, and Hugo A. Hopenhayn. 2006. "A Theory of Financing Constraints and Firm Dynamics." *Q.J.E.* 121:229–65.
Courty, Pascal, and Hao Li. 2000. "Sequential Screening." *Rev. Econ. Studies* 67: 697–717.
DeMarzo, Peter, and Michael Fishman. 2007. "Optimal Long-Term Financial Contracting." *Rev. Financial Studies* 20:2079–2127.
DeMarzo, Peter, and Yuliy Sannikov. 2006. "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." *J. Finance* 61: 2681–2724.
Denis, David, Diane Denis, and Atulya Sarin. 1997. "Ownership Structure and Top Executive Turnover." *J. Financial Econ*. 45:193–221.
Dizdar, Deniz, Alex Gershkov, and Benny Moldovanu. 2011. "Revenue Maximization in the Dynamic Knapsack Problem." *Theoretical Econ.* 6 (2): 157–84.
Edmans, Alex, and Xavier Gabaix. 2009. "Is CEO Pay Really Inefficient? A Survey of New Optimal Contracting Theories." *European Financial Management* 15 (3): 486–96.
———. 2011. "Tractability in Incentive Contracting." *Rev. Financial Studies* 24 (9): 2865–94.
Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov. 2012. "Dynamic Incentive Accounts." *J. Finance* 67 (5): 1603–47.

Eisfeldt, Andrea L., and Adriano A. Rampini. 2008. "Managerial Incentives, Capital Reallocation, and the Business Cycle." *J. Financial Econ.* 87:177–99.

Eso, Peter, and Balazs Szentes. 2007. "Optimal Information Disclosure in Auctions and the Handicap Auction." *Rev. Econ. Studies* 74 (3): 705–31.

Fine, Charles H. 1998. *Clockspeed: Winning Industry Control in the Age of Temporary Advantage*. Cambridge, MA: Perseus.

Fisman, Ray, Rakesh Kuhrana, and Matthew Rhodes-Kropf. 2005. "Governance and CEO Turnover: Do Something or Do the Right Thing?" Manuscript, Columbia Univ. and Harvard Bus. School.

Fong, Yuk-fai, and Jin Li. 2010. "Relational Contracts, Limited Liability, and Employment Dynamics." Manuscript, Northwestern Univ.

Fudenberg, Drew, and David Levine. 1983. "Subgame-Perfect Equilibria of Finite and Infinite Horizon Games." *J. Econ. Theory* 31:251–68.

Garrett, Daniel. 2011. "Durable Goods Sales with Dynamic Arrivals and Changing Values." Manuscript, Northwestern Univ.

Garrett, Daniel, and Alessandro Pavan. 2011*a*. "Dynamic Managerial Compensation: On the Optimality of Seniority-Based Schemes." Manuscript, Northwestern Univ.

———. 2011*b*. "Managerial Turnover in a Changing World." Manuscript, Northwestern Univ.

Gayle, George-Levy, Limor Golan, and Robert Miller. 2008. "Promotion, Turnover and Compensation in the Executive Market." Manuscript, Carnegie Mellon Univ.

Gershkov, Alex, and Benny Moldovanu. 2009*a*. "Dynamic Revenue Maximization with Heterogeneous Objects: A Mechanism Design Approach." *American Econ. J.: Microeconomics* 1 (2): 168–98.

———. 2009*b*. "Learning about the Future and Dynamic Efficiency." *A.E.R.* 99 (4): 1576–87.

———. 2010*a*. "Demand Uncertainty and Dynamic Allocation with Strategic Agents." Manuscript, Univ. Bonn.

———. 2010*b*. "Efficient Sequential Assignment with Incomplete Information." *Games and Econ. Behavior* 68 (1): 144–54.

———. 2010*c*. "Non-Bayesian Optimal Search and Dynamic Implementation." Manuscript, Univ. Bonn.

———. 2012. "Optimal Search, Learning and Implementation." *J. Econ. Theory* 147 (3): 881–909.

Hadlock, Charles J., and Gerald B. Lumer. 1997. "Compensation, Turnover, and Top Management Incentives: Historical Evidence." *J. Bus.* 70:153–87.

He, Zhiguo. 2009. "Optimal Executive Compensation When Firm Size Follows a Geometric Brownian Motion." *Rev. Financial Studies* 22:859–92.

Hermalin, Benjamin, and Michael Weisbach. 1998. "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO." *A.E.R.* 88:96–118.

Holmström, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Rev. Econ. Studies* 66:169–82.

Inderst, Roman, and Holger M. Mueller. 2010. "CEO Replacement under Private Information." *Rev. Financial Studies* 23:2935–69.

Jenter, Dirk, and Katharina Lewellen. 2010. "Performance-Induced CEO Turnover." Manuscript, Grad. School Bus., Stanford Univ.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *J.P.E.* 87: 972–90.

Laffont, Jean-Jacques, and Jean Tirole. 1986. "Using Cost Observation to Regulate Firms." *J.P.E.* 94:614–41.

Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" *J.P.E.* 87:1261–84.
Levitt, Steven D., and Christopher M. Snyder. 1997. "Is No News Bad News? Information Transmission and the Role of 'Early Warning' in the Principal-Agent Model." *Rand J. Econ.* 28:641–61.
McAdams, David. 2011. "Performance Turnover in a Stochastic Partnership." *American Econ. J.: Microeconomics* 3 (4): 107–42.
Milgrom, Paul, and Ilya Segal. 2002. "Envelope Theorems for Arbitrary Choice Sets." *Econometrica* 70:583–601.
Mortensen, Dale T., and Christopher A. Pissarides. 1994. "Job Creation and Job Destruction in the Theory of Unemployment." *Rev. Econ. Studies* 61:387–415.
Mukherjee, Arijit. 2008. "Career Concerns, Matching, and Optimal Disclosure Policy." *Internat. Econ. Rev.* 49:1211–50.
Pai, Mallesh, and Rakesh Vohra. 2011. "Optimal Dynamic Auctions and Simple Index Rules." Manuscript, Northwestern Univ.
Pavan, Alessandro, Ilya Segal, and Juuso Toikka. 2012. "Dynamic Mechanism Design." Manuscript, Massachusetts Inst. Tech., Northwestern Univ., and Stanford Univ.
Rose, Nancy L., and Andrea Shepard. 1997. "Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?" *Rand J. Econ.* 28: 489–514.
Said, Maher. Forthcoming. "Auctions with Dynamic Populations: Efficiency and Revenue Maximization." *J. Econ. Theory.*
Sannikov, Yuliy. 2008. "A Continuous-Time Version of the Principal-Agent Problem." *Rev. Econ. Studies* 75:957–84.
Sen, Arijit. 1996. "Termination Clauses in Long-Term Contracts." *J. Econ. and Management Strategy* 5:473–96.
Shapiro, Carl, and Joseph Stiglitz. 1984. "Equilibrium Unemployment as a Worker Discipline Device." *A.E.R.* 74:433–44.
Shleifer, Andrei, and Robert W. Vishny. 1989. "Management Entrenchment: The Case of Manager-Specific Investments." *J. Financial Econ.* 25:123–39.
Spear, Stephen E., and Cheng Wang. 2005. "When to Fire a CEO: Optimal Termination in Dynamic Contracts." *J. Econ. Theory* 120:239–56.
Stokey, Nancy, and Robert E. Lucas Jr., with Edward C. Prescott. 1989. *Recursive Methods in Economic Dynamics.* Cambridge, MA: Harvard Univ. Press.
Strulovici, Bruno. 2011. "Contracts, Information Persistence, and Renegotiation." Manuscript, Northwestern Univ.
Tchistyi, Alexei. 2006. "Security Design with Correlated Hidden Cash Flows: The Optimality of Performance Pricing." Manuscript, Stern School, New York Univ.
Wang, Cheng. 2011. "Termination of Dynamic Contracts in an Equilibrium Labor Market Model." *J. Econ. Theory* 146 (1): 74–110.
Weisbach, Michael. 1988. "Outside Directors and CEO Turnover." *J. Financial Econ.* 20:431–60.
Williams, Noah. 2011. "Persistent Private Information." *Econometrica* 79 (4): 1233–75.
Yang, Huanxing. Forthcoming. "Nonstationary Relational Contracts." *Internat. Econ. Rev.*
Yermack, David. 2006. "Golden Handshakes: Separation Pay for Retired and Dismissed CEOs." *J. Accounting and Econ.* 41:237–56.
Zhang, Yuzhe. 2009. "Dynamic Contracting with Persistent Shocks." *J. Econ. Theory* 144:635–75.