Threat of Dismissal: Incentive or Sorting?

Illoong Kwon*

University of Michigan

January 2004

Abstract

Many people are fired from their jobs for poor performance. However, it is difficult to distinguish whether they are fired because they are not well suited for their job (sorting explanation) or because the firms are trying to provide incentives for effort (incentive explanation). This paper develops a dynamic incentive model of dismissal and proposes a methodology to distinguish between these two explanations. The methodology rests on the learning-by-doing and the changes in the slope of dismissal probability with respect to tenure. With our unique personnel data, we find significant evidence for the incentive explanation.

JEL Classification: D82, D83, J41, J63

Keywords: Dismissal, moral hazard, sorting

*Correspondence address: Illoong Kwon, Department of Economics, University of Michigan, 611 Tappan St., Ann Arbor, MI 48109, USA; email: ilkwon@umich.edu. I would like to thank Charlie Brown, Robert Gibbons, Michael Gibbs, Gary Solon and seminar participants at Washington University, Michigan State University, and NBER personnel economics meeting for helpful comments and suggestions. I wish to thank late Zvi Griliches who gave me the access to the dataset. I am also grateful to Kyoungsoo Yoon for excellent research assistance. All remaining errors are mine.

1 Introduction

Many people are fired from their jobs for poor performance. Their jobs can range from a President to a football coach, a CEO to a salesman. Despite the social and economic importance of these dismissals, there is little understanding about whether these workers are fired because they are not well suited for their jobs or because the principal (or firms) is trying to provide incentives for effort. These are two very different reasons for dismissals. Dismissals that are intended to 'sort out' the bad type (in terms of matching quality or ability, for example) are socially and personally costly. For firms, not only there are search and training costs of new employees after dismissals, but also there are foregone profits from bad matching. Similarly, employees have to search for a new job and suffer from unemployment in the mean time. Dismissals that are to provide 'incentives' are also costly. In addition to the search cost, to provide incentives, firms may have to fire experienced and productive employees if they have poor performance due to some random events. Otherwise, firms would lose commitment power. The remedies for these costs are also quite different depending on the reasons. If most dismissals are due to the 'sorting', then education, job-training, or screening are important in reducing such costs. If most dismissals are due to the 'incentive' reasons, then monitoring or better performance measures are more important in reducing such costs.

Despite these differences, there are few studies that recognize these two alternative explanations for dismissals and even fewer that try to distinguish between these two. This paper develops a dynamic incentive model of dismissal and provides a methodology to distinguish these two models. Also using unique personnel records, we also provide evidence for an incentive model.

These two explanations for dismissals, that is, the sorting explanation and the incentive explanation, are hard to distinguish empirically. Both explanations predict that the probability of dismissal will decrease with an agent's performance. In a sorting model, good performance increases the principal's belief in the agent's type and reduces the dismissal probability. In an incentive model, good performance increases the likelihood of the agent's high effort and reduces the dismissal probability. There are many empirical studies that show this negative relationship between the probability of dismissal and performance (see Weisbach (1988), Jensen and Murphy (1990), Kaplan (1994), Denis and Denis (1995), Conyon (1998), and Chevalier and Ellison (1999)). However, some studies interpret their findings as evidence of an incentive model, and some interpret them as evidence of a sorting model. To our knowledge, no previous study has attempted to distinguish an incentive explanation from a sorting explanation for the threat of dismissal. This is surprising because the distinction between 'hidden action' and 'hidden type' and the difficulty of distinguishing two are well known in the insurance literature. (see e.g. Abbring, Chiappori, and Pinquet (2003))

We present a dynamic incentive model and a sorting model of dismissal, and characterizes the optimal dismissal policy in each model. Both models predict that the dismissal probability decreases with performance and that with the agent's learning-by-doing, the average dismissal probability decreases over time. We show, however, that the *slope* of the dismissal probability increases (in absolute value) over time under an incentive model, but decreases under a sorting model. This difference in the dynamic property of the dismissal policy provides an empirical methodology to distinguish between these two models.

Intuitively, under an incentive model, a principal can provide incentives with the threat of dismissal in two ways. First, she can increase the *level* of the dismissal probability for a given performance. Second, she can also increase the *slope* (in absolute value) of the dismissal probability with respect to performance. That is, the principal can make the dismissal probability more sensitive to performance. Through learning-by-doing, an existing agent becomes more productive than a new one, and as an agent's tenure increases, the agent becomes more expensive to dismiss. Therefore, the principal must decrease the level of dismissal probability. Then, to compensate for the loss of incentives, the principal will increase the slope (in absolute value) of the dismissal probability.

In a sorting model, as the agent's tenure increases, the principal receives more observations on the agent's performance. Then, a new observation has a smaller and smaller effect on the principal's belief about the agent's type (or ability) and also on the dismissal probability. Therefore, the dismissal probability becomes less sensitive to performance as the agent's tenure increases.

Using unique personnel records from a large company in the U.S., we also provide an empirical analysis of dismissals. As both models predict, the dismissal probability decreases with the worker's performance, and the average dismissal probability also decreases with the agent's tenure. However, the slope of the dismissal probability increases (in absolute value) with the agent's tenure, which is consistent with an incentive model.

The incentive model is based on a standard dynamic moral hazard model with limited liability. Our model departs from the standard model by allowing the principal to fire the agent at the end of each period based on his performance. We show that it is optimal to use both the threat of dismissal and the wage contract as incentives devices. This result is not obvious because the principal can lower the wages instead of firing the agent. For example, when the agent's worst payoff upon dismissal is zero, the principal can also provide zero wages instead of firing him. Under the limited liability, however, we show that without the dismissal, the principal cannot provide zero *continuation* payoffs and satisfy the incentive constraint at the same time, while the principal can do both if she uses the threat of dismissal.

While there is an extensive literature on dynamic incentive contracts, most studies have focused on the wage contract only (e.g. Rogerson (1985)). On the other hand, most previous studies on the threat of dismissal typically assume that the wage cannot be contingent on the agent's performance (see Shapiro and Stiglitz (1984), Calvo (1985), Kuhn (1986), Sparks (1986), and Mori (1998)). However, it is not clear why the dismissal decision can be contingent on the performance, but the wage contract cannot. Furthermore, if we allow for a performance-based wage contract in these models, using the threat of dismissal in these models is typically not optimal (see Kwon (2003)), even though most firms often use both the wage and the threat of dismissal as incentive devices. There is also a large literature on dynamic learning and matching (e.g. Jovanovic and Nyarko (1996)). However, few have analyzed the slope of the dismissal (or separation) probability which is the key in distinguishing the incentive explanation and the sorting explanation.

The paper is organized as follows. Section 2 provides a simple sorting model as a benchmark. Then, in section 3, we introduce a dynamic incentive model of dismissal and characterize the optimal contract. Section 4 provides an empirical analysis of dismissal using a unique personnel records of health insurance claim processors in a large US insurance company. Section 5 concludes.

2 Sorting Model

For a benchmark, we first present a multi-period sorting model in which a principal can fire an agent with positive probability when she believes that the agent's type is below a certain threshold. That is, the dismissal serves to 'sort out' the bad-type workers.

Assume that the agent's type θ is unknown to both the principal and the agent. For simplicity, we assume that θ is a *firm-specific* matching quality that both the principal and the agent do not observe ex-ante. We also normalize the outside wage to zero. Thus, the wages will stay at zero regardless of the history of outcomes. As we show in the appendix, however, we can easily accommodate the general ability in the model and analyze the wages more explicitly without changing the qualitative results of the analysis.

Suppose that an agent' performance in period t (y_t) is given by

$$y_t = \theta_t + \epsilon_t \tag{1}$$

where θ_t is the agent's type and ϵ_t is random noise ($t = 1, 2, 3...$).

To compare with an incentive model in next section, we will allow learning-by-doing. That is,

$$\theta_t = \theta_0 + k_t \tag{2}$$

where k_t is the human capital of an agent. Through learning-by-doing, k_t increases deterministically over time, that is, $k_t = k_{t-1} + l$ where l (≥ 0) is the amount of learning-by-doing in each period.[1] Without loss of generality, assume $k_1 = 0$.

The principal observes y_t and k_t, but neither the principal nor the agent observes θ_0 or ϵ_t. The prior distribution of θ_0 follows a normal distribution, $N(0, \sigma_1^2)$. ϵ_t's are independently and identically distributed according to a normal distribution $N(0, \sigma_\epsilon^2)$. Denote the posterior distribution of y_t conditional on the history of the agent's performance $H_{t-1} = (y_1, ..y_{t-1})$ by

$$y_t | H_{t-1} \sim N(m_t, \sigma_t^2) \tag{3}$$

It is well known that for a given y_t, the posterior distribution for y_{t+1} is

$$y_{t+1} | (y_t, H_{t-1}) \sim N(m_{t+1}, \sigma_{t+1}^2) = N(\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l, \frac{\sigma_t^2 \sigma_\epsilon^2}{\sigma_\epsilon^2 + \sigma_t^2}) \tag{4}$$

where l is the amount of learning-by-doing in period t.

At the beginning of each period, the principal decides whether to renew the contract or to terminate the contract based on the history of the agent's performance. Since (m_t, t) are sufficient statistics[2] for H_{t-1}, denote the probability of termination (or dismissal) at the beginning of period $t + 1$ by $q_t(y_t; m_t)$.

Throughout the paper, we assume that the principal needs to maintain a constant number, N, of employees. Therefore, if a principal dismisses n agents, she must hire n new ones[3]. This

[1] Even if l decreases over time, the qualitative results of our model do not change.

[2] From (4), (m_t, σ_t^2) are sufficient statistics for H_{t-1}. Since σ_t^2 evoles deterministically over time, we can also say that (m_t, t) are sufficient statistics for H_{t-1}.

[3] We assume that the principal can not re-hire an old agent who she has fired before.

assumption is also consistent with the hiring pattern of the company that we analyze in the empirical section of this paper, where the new hirings are mainly done to fill out the vacancies. We assume that there is a cost associated with firing and hiring. In particular, we assume the following:

Assumption 1 The marginal cost of firing and hiring increases in n.

This assumption tells us that finding ten good managers is typically much more difficult and costly than finding one good manager. In particular, we assume that the cost of hiring n new employees is $\frac{\phi n^2}{2}$ where $n = Nq$ and $\phi > 0$.[4] Then, the average cost per employee is $\frac{1}{N}\frac{\phi(Nq)^2}{2} = \frac{\phi N}{2}q^2$. This functional form is for simplicity only and is not responsible for the main qualitative results. The hiring cost includes, among other things, the foregone profits during the search as well as the search and training costs. This marginal forgone profit, that is, the marginal opportunity cost of the firing, increases with n in general. Also, while there can be economies of scale for small n, for n large enough, the marginal search and training cost also increase. Technically, this assumption also provides an interior solution for q_t.

Like Jovanovic and Nyarko (1996), in this section we assume, for simplicity, that the principal is *myopic*: she maximizes the current period payoff in each period. Then, the principal's optimization problem at the beginning of period $t+1$ is:

$$\max_{0 \leq q_t(y_t; m_t)) \leq 1} E[\pi_{t+1}|y_t, m_t] = q_t(y_t; m_t)0 + (1 - q_t(y_t; m_t)) \left(\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l \right) - \frac{\phi N}{2} q_t(y_t; m_t)^2 \tag{5}$$

Note that if the principal fires an agent and hire a new one, the new agent's expected ability is zero since he has no experience and the prior belief on θ_0 is zero. Then, for an interior solution,

[4] That is, N is large enough that for the principal n is given deterministically by Nq.

the FOC for $q_t(y_t; m_t)$ is:

$$-\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} - l - \phi N q_t(y_t; m_t) = 0 \tag{6}$$

Therefore, the optimal dismissal policy is:

$$q_t^*(y_t; m_t) = \begin{cases} 1 & \text{if } y_t < y_t^1(m_t) \\ -\frac{1}{\phi N}\left(\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l\right) & \text{if } y_t^1(m_t) \leq y_t \leq y_t^0(m_t) \\ 0 & \text{if } y_t > y_t^0(m_t) \end{cases} \tag{7}$$

where $y_t^1(m_t) = -\frac{\sigma_\epsilon^2 m_t + \phi N(\sigma_\epsilon^2 + \sigma_t^2) + l(\sigma_\epsilon^2 + \sigma_t^2)}{\sigma_t^2}$ and $y_t^0(m_t) = -\frac{\sigma_\epsilon^2 m_t + l(\sigma_\epsilon^2 + \sigma_t^2)}{\sigma_t^2}$.

This optimal dismissal policy has the following features:

Proposition 1 *(i) For* $0 < q_t^*(y, m) < 1$, $\frac{\partial q_t^*(y;m)}{\partial y} < 0$, *and* $\frac{\partial^2 q_t^*(y,m)}{\partial t \partial y} > 0$.

(ii) For $m > 0$, *both* $q_t^*(y; m)$ *and* $E_{y_t}[q_t^*(y_t; m)]$ *(weakly) decrease with* t.

Proof. See appendix. ■

Poor performance therefore leads to an unfavorable update in the principal's belief about the agent's ability, and increases the probability of dismissal. Furthermore, regardless of the history of outcomes, the *slope* of the dismissal probability becomes flatter over time. That is, the dismissal probability becomes less sensitive to performance over time. This is the case because the informational value of additional observation decreases with the number of observations. This result does not depend on learning-by-doing because the slope ($= -\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_\epsilon^2 + \sigma_t^2}$) does not depend on l. As we show in the next section, this is an important feature that will allow us to distinguish a sorting model from an incentive model.

Because of learning-by-doing, m_t is most likely to be positive. As σ_t^2 decreases over time, the dismissal probability also decreases for a given y_t. Since, on average, m_t increases over time for those who remain, the dismissal probability (both for a given y and on average) should decreases even further.

Figure 1 illustrates the change in the optimal dismissal rule over time.

[Figure 1 here]

3 Incentive Model

In this section, we develop an incentive model of dismissal.[5] Firms typically use both the threat of dismissal and the wage contract as incentive devices. However, as we discussed above, most previous theoretical models do not incorporate both. We are interested in when it is optimal to use both devices and particularly what the features of the dismissal policy are that will allow us to distinguish it from a sorting model.

3.1 Setup

Consider a standard moral hazard model with two effort levels and continuous outcomes. There are two periods, period $t = 1$ and period $t = 2$. We will later extend the model to an infinite horizon model. In each period, the principal's revenue (y_t) is given by

$$y_t = \theta_t + a_t + \epsilon_t \tag{8}$$

where θ_t is the agent's human capital in period t, a_t is the agent's effort level, and ϵ_t is i.i.d. random noise. In each period, an agent chooses one of two effort levels, $a_t \in \{a_H, a_L\}$ where $a_H > a_L$.

The agent's human capital increases with his tenure at the firm through learning-by-doing following (2)[6]. Without loss of generality, let $\theta_0 = 0$. Suppose that the agent already has τ

[5] Even thoiugh a sorting model and an incentive model are not mutually exclusive, we focus on the incentive model to highlight the main features of the incentive explanation better.

[6] In this model, human capital increases deterministically, regardless of the agent's effort choice. Kwon (1999)

periods of tenure. Then, $\theta_1 = k_\tau = \tau l$. We assume that this human capital is firm-specific, so it does not affect the agent's reservation utility.

Unlike the sorting model, the principal observes θ_t as well as y_t. However, she does not observe a_t nor ϵ_t. Denote the probability density function of y_t given θ_t and a_t by $f(y_t|\theta_t, a_t)$. We assume that $\frac{f(y_t|\theta_t,a_H)}{f(y_t|\theta_t,a_L)}$ is strictly increasing in y_t for any θ_t (monotone likelihood ratio property, hereinafter MLRP). In addition, we assume the following:

Assumption 2 For any θ_t, $\lim_{y_t \to \infty} \frac{f(y_t|\theta_t,a_H)}{f(y_t|\theta_t,a_L)} = \infty$ and $\lim_{y_t \to -\infty} \frac{f(y_t|\theta_t,a_H)}{f(y_t|\theta_t,a_L)} = 0$.

This assumption can be relaxed at the cost of more notation but without additional new insights. Thus, we will maintain this assumption throughout the paper. This assumption holds if ϵ follows a normal distribution.

The principal is risk-neutral while the agent is risk-averse. The agent has a separable and additive utility function, $\sum_t \delta^{t-1}[u(w_t) - g(a_t)]$ where $u(0) = 0, u' > 0$ and $u'' < 0$. Both the principal and the agent have the same discount factor δ $(0 \leq \delta < 1)$. w_t is the wage payment in period t, and $g(a_t)$ is the disutility of effort. Denote $g(a_H)$ by H (> 0) and let $g(a_L) = 0$[7]. Also for simplicity, assume that the agent's reservation utility is zero[8]. In addition, we make the following assumption:

Assumption 3 The agent has *limited liability*, that is, $w_t \geq 0$ for all t.

analyzes the optimal wage contract (but not the threat of dismissal) when human capital increases only when the agent chooses a high effort.

[7]As long as $g(a_H) - g(a_L)$ is positive and large enough, the qualitative results of the analysis do not change even if $g(a_L) \neq 0$.

[8]Since we have normalized $u(0) = 0$, this is not a normalization. For our analysis, it is sufficient if the reservation utility is low enough that the participation constraint is not binding with the limited liability constraint, which is true if $g(a_H) - g(a_L)$ is large enough.

The principal cannot commit to a long-term contract. So, in each period, the contract is determined in the spot market. Thus, the optimal contract will not depend on the history of the agent's previous performance. These assumptions are for simplicity, and allowing for a long-term contract does not change our qualitative results.

An important deviation from a standard moral hazard model is that in addition to a wage contract, $w_t(y_t)$, we allow the principal to fire an agent at the end of each period t based on the outcome y_t. Denote the probability of dismissal at the end of period t by $q_t(y_t)$ where $0 \leq q_t(y_t) \leq 1$. We maintain assumption 1. So the marginal cost of dismissal increases with q.

The timing of the game is as follows: At the beginning of period t, the principal and an agent write a (short-term) contract specifying $w_t(y_t)$ and $q_t(y_t)$. Then, the agent chooses an effort level $a_t \in \{a_H, a_L\}$. At the end of each period, the outcome (y_t) is realized, and the payments and the dismissal decisions are made according to the contract. If the principal fires an agent, she hires a new one before the next period begins.

3.2 Optimal Contract

The optimal contract has two components: the wage contract $w_t(y_t)$ and the dismissal rule $q_t(y_t)$.[9] This section explores whether it is optimal to use the threat of dismissal as an incentive device in the first period. The dismissal is obviously a punishment tool that provides a zero continuation payoff to an agent. However, notice that the wage contract can also provide a zero continuation payoff by offering zero wages in the second period regardless of the second period outcome. Given that the wage contract can provide the same amount of punishment as the dismissal, a puzzle is why a principal would use the costly threat of dismissal.

[9] In principle, the wage contract $w_t(y_t)$ can be different depending on whether the agent gets dismissed or not. However, it is straightforward to show that the optimal wage contract does not depend on the dismissal decision because it only provides unnecessary risks to the agent.

3.2.1 The Second Period

We solve the game backwards. Since the dismissal decision does not matter in the second period, we only characterize the optimal wage contract. It is often more convenient to consider the wage contract in terms of *utils*. Define $v_t(y_t) \equiv u(w_t(y_t))$ or $w_t(y_t) \equiv h(v_t(y_t))$ where $h = u^{-1}$. The principal's optimization problem in the second period is:

$$\max_{v_2(y_2)} \int [y_2 - h(v_2(y_2))] f(y_2; \theta_2, a_H) dy_2$$

subject to

$$\int v_2(y_2) f(y_2; \theta_2, a_H) dy_2 - H \geq 0 \tag{9}$$

$$\int v_2(y_2) f(y_2; \theta_2, a_H) dy_2 - H \geq \int v_2(y_2) f(y_2; \theta_2, a_L) dy_2 \tag{10}$$

$$v_2(y_2) \geq 0 \text{ for } \forall y_2 \tag{11}$$

where (9) is the participation constraint, (10) is the incentive constraint, and (11) is the limited liability constraint.

Lemma 1 *The participation constraint (9) holds with strict inequality.*

Proof. See appendix. ■

Because of the limited liability constraint, the principal cannot implement a large punishment (i.e. negative wages) for low performance, and the principal must provide incentives by rewarding good performance. Therefore, the participation constraint is not binding.

For any given y_2, the FOC for $v_2(y_2)$ is

$$\frac{\partial \mathcal{L}}{\partial v_2(y_2)} = -h'(v_2(y_2)) f(y_2; \theta_2, a_H) + \lambda_2 \left(f(y_2; \theta_2, a_H) - f(y_2; \theta_2, a_L) \right) + \gamma_2(y_2) = 0 \tag{12}$$

where λ_2 and $\gamma_2(y_2)$ are the Lagrange multipliers for (10) and (11), respectively. It is easy to see that the incentive constraint is binding and $\lambda_2 > 0$. Therefore,

$$v_2^*(y_2;\theta_2) = g(\lambda_2(1 - \frac{f(y_2;\theta_2,a_L)}{f(y_2;\theta_2,a_H)}) + \frac{\gamma_2(y_2)}{f(y_2;\theta_2,a_H)}) \tag{13}$$

where $g \equiv h'^{-1}$. Note that $g' > 0$ since $h'' > 0$.

Define $\overline{y}_2(\theta_2)$ such that $g(\lambda_2(1 - \frac{f(\overline{y}_2;\theta_2,a_L)}{f(\overline{y}_2;\theta_2,a_H)})) = 0$. From MLRP, if $y_2 > \overline{y}_2(\theta_2)$, $g(\lambda_2(1 - \frac{f(y_2;\theta_2,a_L)}{f(y_2;\theta_2,a_H)})) > 0$ and $\gamma_2(y_2) = 0$. Therefore, the optimal contract $v_2^*(y_2;\theta_2)$ should take the following form:

$$v_2^*(y_2;\theta_2) = \begin{cases} g(\lambda_2(1 - \frac{f(y_2;\theta_2,a_L)}{f(y_2;\theta_2,a_H)})) & \text{if } y > \overline{y}_2(\theta_2) \\ 0 & \text{if } y \leq \overline{y}_2(\theta_2) \end{cases} \tag{14}$$

Thus, the principal provides zero wages if the performance is below $\overline{y}_2(\theta_2)$. If the performance exceeds $\overline{y}_2(\theta_2)$, then the wages increase in performance. Often $\overline{y}_2(\theta_2)$ is called a 'target performance' level where a bonus is promised if the performance exceeds this target.

To analyze the dismissal decision in the first-period, we need to look at both the principal's and the agent's continuation payoff. Thus, define the agent's expected utility in the second period by $V_2^* \equiv \int v_2^*(y_2;\theta_2)f(y_2;\theta_2,a_H)dy_2 - H$. Then, the following lemma holds:

Lemma 2 *V_2^* does not depend on θ_2.*

Proof. See appendix. ■

Intuitively, because θ_2 is known to the principal, it does not provide any informational rent to the agent. It only shifts both the probability distribution function of y_2 and the wage function $v_2^*(y_2;0)$ to the right by the same amount, θ_2. Thus, the expected wage payment in the second period does not change with θ_2.

As expected revenue increases with θ_2, however, the principal's profit in the second period increases with θ_2. Denote the principal's second period profit by $\Pi_2^*(\theta_2)$. Then,

Lemma 3 $\Pi_2^*(\theta_2)$ *increases with* θ_2.

Proof. Follows from lemma 2. ■

3.2.2 The First Period

Suppose that the agent already has τ periods of tenure at the firm, i.e., $\theta_1 = k_\tau$. If the agent does not get fired, then $\theta_2 = k_{\tau+1}$. But, if the principal fires the agent and hires a new one, $\theta_2 = k_0$. Therefore, in the first period, the principal's optimization is as follows:

$$\max_{q_1(y_1), v_1(y_1)} \int \left[y_1 - h(v_1(y_1)) + q_1(y_1)\delta\Pi_2^*(k_0) + (1 - q_1(y_1))\delta\Pi_2^*(k_{\tau+1}) - \frac{\phi N}{2}(q_1(y_1))^2 \right] f(y_1|\theta_1, a_H)dy_1$$

subject to

$$\int [v_1(y_1) + (1 - q_1(y_1))\delta V_2^*] f(y_1|\theta_1, a_H)dy_1 - H \geq 0 \tag{15}$$

$$\int [v_1(y_1) + (1 - q_1(y_1))\delta V_2^*] f(y_1|\theta_1, a_H)dy_1 - H \tag{16}$$

$$\geq \int [v_1(y_1) + (1 - q_1(y_1))\delta V_2^*] f(y_1|\theta_1, a_L)dy_1$$

$$v_1(y_1) \geq 0 \text{ for } \forall y_1 \tag{17}$$

$$0 \leq q_1(y_1) \leq 1 \text{ for } \forall y_1 \tag{18}$$

where (15) is the participation constraint and (16) is the incentive constraint.

The FOCs for $q_1(y_1)$ and $v_1(y_1)$ yield the following proposition:

Proposition 2 *For any given* $\theta_1 (= k_\tau)$, *there exist* $\overline{y}_1(\theta_1)$, $y^1(\theta_1)$, *and* $y^0(\theta_1)$ *such that* $y^1(\theta_1) <$

$y^0(\theta_1) < \overline{y}_1(\theta_1)$ and the first-period optimal contract takes the following form:

$$v_1^*(y_1;\theta_1) = \begin{cases} g(\lambda_1(\theta_1)(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)})) & \text{if } y \geq \overline{y}_1(\theta_1) \\ 0 & \text{if } y < \overline{y}_1(\theta_1) \end{cases} \tag{19}$$

$$q_1^*(y_1;\theta_1) = \begin{cases} 1 & \text{if } y_1 \leq y^1(\theta_1) \\ \frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)} - 1) - \Delta_{\tau+1}(\theta_1)] & \text{if } y^1(\theta_1) < y_1 \leq y^0(\theta_1) \\ 0 & \text{if } y_1 > y^0(\theta_1) \end{cases} \tag{20}$$

where $\lambda_1(\theta_1)$ is the Lagrange multiplier for the first period incentive constraint (16), *and* $\Delta_{\tau+1}(\theta_1) \equiv \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$.

Proof. See appendix. ■

Because $\frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)}$ is decreasing in y_1, the dismissal probability also decreases in y_1. In other words, poor performance leads to a higher probability of dismissal. Thus, it is optimal to use the threat of dismissal as well as the wage contract as incentive devices. Also notice that $y^0(\theta_1) < \overline{y}_1(\theta_1)$. So the principal does not provide a positive wage and fire the agent. See Figure 2.

[Figure 2]

In order to understand why a principal should use the costly threat of dismissal as an incentive device, let us consider the costs and benefits of using the threat of dismissal. There are two kinds of costs associated with the dismissal. First, there is a new hiring cost after the dismissal of an old agent, $\frac{\phi N}{2}(q_1(y_1))^2$. Second, there is a loss of profit because the new agent has less human capital, $\Delta_{\tau+1}(\theta_1) \equiv \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$.

The benefit of using the dismissal is that *along the equilibrium path* the dismissal provides a larger punishment for a bad first period performance than the wage contract alone. In the first period, the worst *feasible* punishment for a bad performance is to receive a zero wage in the first period *and* also a zero (continuation) payoff in the second period. In principle, both the wage contract and the dismissal can provide zero payoffs in the second period. That is, the principal can provide zero wages in the second period regardless of the outcome. Even though providing zero wages in the second period (regardless of the second period outcome) is feasible, it is not optimal since it violates the second period incentive constraint. In order to satisfy the second period incentive constraint, from lemma 1, the principal should give a positive net payoff to the agent. Therefore, the wage contract alone cannot provide a zero second period payoff to the agent in the first period along the equilibrium path.

On the other hand, if the principal fires the agent in the first period, the second period surplus V_2^* goes to a new agent, so it does not provide a positive continuation payoff to the agent in the first period. Therefore, the benefit of the dismissal is that it can provide the largest feasible punishment to the agent in the first period without violating the second period incentive constraint.

When the first period's performance is poor and requires a large punishment, the benefit of using a dismissal outweighs the cost. So it is optimal to fire an agent with a positive probability if his performance is bad enough. Therefore, the threat of dismissal is an optimal incentive device even when the principal uses a performance-based wage contract. Note that this result arises naturally in any repeated moral hazard model with a limited liability assumption if we allow the principal to fire the agent. We do not need any exogenously given benefit of the job or a negative payoff (such as humiliation) from the dismissal to explain the threat of dismissal as an incentive device.

3.3 Tenure and Optimal Contract

This section studies how the optimal contract changes with the agent's tenure τ in the first period. Because the principal and the agent sign a contract in the spot market every period, the agent's tenure affects the optimal contract only through his human capital $\theta_1 = k_\tau$. We can also denote the optimal contract by $v_1^*(y_1;\tau)$ and $q_1^*(y_1;\tau)$.

Suppose that the agent's tenure in the first period ,τ, becomes larger. Then both k_τ and $k_{\tau+1}$ increase. This increase has three effects on the optimal contract:

First, from (8), the increase in τ (or $\theta_1 = k_\tau$) shifts the probability density functions, $f(y_1|\theta_1, a_H)$ and $f(y_1|\theta_1, a_L)$, to the right. Thus, the optimal contract (19) and (20) also shift to the right by the same amount. Notice, however, that since both the contract and the probability density function shift by the same amount, this effect does not change the expected wage payment or the expected dismissal probability. In particular, this effect does not change either the participation constraint or the incentive constraint. Therefore, it is often more convenient to consider the contract in terms of $\tilde{y}_t$ where

$$\tilde{y}_t \equiv y_t - k_t \tag{21}$$

Then, we can ignore this first effect on the optimal contract if we write the contract in terms of $\tilde{y}_t$, i.e. , $v_1^*(\tilde{y}_1;\tau)$ and $q_1^*(\tilde{y}_1;\tau)$. We can interpret $\tilde{y}_t$ as human capital adjusted performance.

Second, the increase in τ also raises $\Delta_{\tau+1} = \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$ from lemma 3. From (20), this increase causes the level of dismissal probability to decrease.

Third, we can show that the slopes of the dismissal probability and the wage contract increase (in absolute value) with the agent's tenure τ. This effect is not immediately obvious and it is stated in the following proposition.

Proposition 3 *(i) For* $0 < q_1(\tilde{y}_1;\tau) < 1$, $\frac{\partial^2 q_1(\tilde{y}_1;\tau)}{\partial\tau\partial\tilde{y}_1} < 0$.

(ii) For $v_1(\tilde{y}_1;\tau) > 0$, $\frac{\partial^2 v_1(\tilde{y}_1;\tau)}{\partial\tau\partial\tilde{y}_1} > 0$ *and* $\frac{\partial v_1(\tilde{y}_1;\tau)}{\partial\tau} > 0$.

(iii) $\frac{\partial E[q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau} < 0$ *and* $\frac{\partial E[v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau} > 0$.

Proof. See appendix. ■

Therefore, both the dismissal probability and wage contract become more sensitive to the performance as the tenure increases. Furthermore, the average dismissal probability decreases with tenure, but the average wage increases.

To understand the intuition behind this result, note that the threat of dismissal, $q_1^*(\tilde{y}_1;\tau)$, affects the agent's incentives in two different ways. First, the threat of dismissal can provide stronger incentives by increasing the *level* of the dismissal probability (or shifting y^1 and y^0 to the right), because it decreases the expected payoffs for bad performance[10]. Second, it can provide stronger incentives by increasing the *slope* (in absolute value) of the dismissal probability.

As τ increases, the agent's human capital increases and becomes more expensive to replace, i.e., $\Delta_{\tau+1}(\theta_1) \equiv \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$ increases. Thus, in order to reduce the number of employees fired ($n = Nq$), the principal should decrease the *level* of the dismissal probability. However, this reduces the incentives and violates the incentive constraint. So in order to compensate for the loss of incentives, the principal must increase the *slope* (in absolute value) of the dismissal probability as well as the slope of the wage function. See Figure 3(a).

[Figure 3 here]

As Figure 3(a) shows that for any $\tilde{y}_1$, the wage increases with tenure. Therefore, the expected wage will also increase with tenure. The upward sloping (or backloaded) wage profile over tenure

[10] To be more precise, we also need to prove that the dismissal probability is positive only for the performance where the likelihood of bad outcome is larger than the likelihood of good outcome. See the proof of proposition 2 for details.

has been widely observed in the literature. While there are many different explanations for this pattern, most of them do not apply to a short-term contract with firm-specific human capital. Our model shows, however, that once we allow for the threat of dismissal as an incentive device, then the (average) wage increases with tenure even under the short-term contract and with firm-specific human capital.

On the other hand, as Figure 3(a) shows, the change in the dismissal probability is somewhat ambiguous. Because the slope is increasing, it is possible that the dismissal probability may increase for $\tilde{y}$ near y^1.

Even when we write the contract in terms of y_t $(= \tilde{y}_t + k_t)$, Proposition 3 still holds in that the slopes of both the dismissal probability and the wage function increase (in absolute value) with tenure τ, as shown in Figure 3(b). Since $E_{\tilde{y}_1}[v_1(\tilde{y}_1;\tau)] = E_{y_1}[v_1(y_1;\tau)]$, the expected wage in terms of y_t also increases with tenure.

The increase in the slope (in absolute value) of the dismissal probability is particularly interesting because the sorting model predicts the opposite as shown in Proposition 1. An astute reader will notice, however, that Proposition 3 is not directly comparable to Proposition 1. The exercise in this section is like comparing the optimal contracts of two agents with different levels of tenure when both have only one more period in the contract remaining. Thus, Proposition 3 does not directly applies to the change of the optimal contract of one agent over his tenure. Thus, next section extends this result to an infinite horizon model.

3.4 Infinite Horizon Model and Slope of Dismissal Probability

As we show below, even in the infinite horizon model, the slope of the dismissal probability increases (in absolute value) with an agent's tenure due to the exact same intuitions as in the two-period model. Therefore, in this section, we only make several technical points in making extension to an infinite horizon model.

First, note that the principal's optimization in any period t is identical to the first period problem of the two period model. We just need change the subscript 1 to t and 2 to $t+1$, and re-interpret Π^*_{t+1} and V^*_{t+1} as the continuation payoffs for the principal and the agent, respectively. Thus, proposition 2 immediately extends to an infinite horizon model.

From the incentive constraint and the limited liability condition, lemma 1 also immediately extends to an infinite horizon model. It is also clear that the continuation profit of the principal increases with the agent's human capital. Thus, lemma 3 still holds, that is, $\Delta_{\tau+1}$ increases with the agent's tenure τ. On the other hand, lemma 2 no longer holds because the dismissal probability changes with tenure, the agent's continuation payoff will also change with his tenure (or θ_t). In fact, proposition 3(iii) suggests that as the agent's tenure increases, the expected wage payment increases and the expected dismissal probability decreases. Therefore, we can show that the continuation payoff increases with the agent's tenure.

Lemma 4 *The agent's continuation payoff, V_t^*, increases with his tenure.*

Proof. See appendix. ■

From (20), the increase in V_t^* also raises the slope of the dismissal probability (in absolute value). The following proposition shows that even after considering the adjustment of λ_t, the slope is increasing (in absolute value) with the agent's tenure.

Proposition 4 *For* $0 < q(\tilde{y}; t) < 1$, $\frac{\partial^2 q(\tilde{y};t)}{\partial t \partial \tilde{y}} < 0$.

Proof. See appendix. ■

Thus, even in the infinite horizon model, the slope of the dismissal probability increases (in absolute value) with an agent's tenure. Note that the incentive model and the sorting model provide opposite predictions on the change in the slope of dismissal probability with the agent's tenure.

4 Empirical Analysis

In this section, using unique personnel data from a large US insurance company, we investigate the use of the threat of dismissal in this firm.

In previous sections, we have analyzed the threat of dismissal in a sorting model and also in an incentive model. Even though the two models are very different, it is empirically difficult to distinguish between the two. In both models, the dismissal probability is decreasing in the agent's performance, and the average dismissal probability decreases with the agent's tenure. Perhaps because of this, few previous studies have recognized these two alternative explanations for the threat of dismissal, and none has attempted to distinguish between these two models. That is, there is little quantitative evidence for the use of threat of dismissal as an incentive device.

Our model shows, however, that with learning-by-doing, the two models make the opposite predictions about the changes in the *slope* of the dismissal probability. In a sorting model, the slope of the dismissal probability decreases (in absolute value) with the agent's tenure, while in an incentive model, the slope increases with the agent's tenure. Therefore, in principle, we can distinguish these two models empirically based on the change in the slope of the dismissal probability with respect to the tenure. Of course, in reality, a firm is likely to face both the sorting and incentive problems. The change in the slope of the dismissal probability will at least tell us which model dominates in data. If the slope increases (in absolute value) with the agent's tenure, it will provide evidence for the use of threat of dismissal as in an incentive model.

Table 1 summarizes the predictions of the sorting model and the incentive model. Note that as Figure 3(b) shows, the dismissal probability for a given performance may increase or decrease with tenure under the incentive model, because the probability distribution of performance changes from learning-by-doing.

[Table 1 here]

4.1 Data

The data are from the personnel records of insurance claim processors in a large insurance company in the U.S.. The original dataset includes 5,888 processors over a two and a half year period (01/01/93-06/30/95). Of this group, we restrict the focus to 3,231 full-time employees working only on indemnity claims.[11] The data contain detailed information on employee performance, compensation, and termination. Table 2 reports some summary statistics.

This dataset is ideal for our analysis because the employees' tasks are essentially the same cross employees and cross different job-levels[12]. Also, they do not get transferred (or promoted) to (and from) other jobs in the company.[13] That is, these workers comprise an well-defined homogenous internal labor market. Therefore, we can ignore the other parts of the company in the analysis and avoid the difficult task of accommodating different qualities of the jobs cross workers. Furthermore, the data contain *objective and consistent* performance measures which is essential for the analysis as we discuss below.

[Table 2 here]

[11] The rest of the processors work on HMO claims. From a workplace perspective, the nature of HMO claims processed at this company appears to be sufficiently different from that of indemnity claims. Less than 0.5% of processors work on both indemnity and HMO claims. These processors are excluded.

[12] If the tasks are qualitatively different cross different job-levels, then under the sorting explanation, we can expect the increased turnovers right after the job changes. However, the data shows no such pattern. This result is not reported in the paper, but available from the author.

[13] See Kwon (1999) for details.

4.1.1 Demographics

About 90% of the employees are female, and 56% of them are married. The average age is 31 years old. Most of employees have a high school diploma, and about 30% of them have a college education or higher. On the whole, these employees can be characterized as female, white-collar, non-managerial, service industry, full-time workers. Even though this group of employees is growing fast in the economy, few studies have examined them.

4.1.2 Measurement of Performance and Compensation

Performance is measured by the *weighted* number of claims processed a day. The company has developed a weighting system to reflect the different types of claims within and cross different job-levels[14]. This measure provides not only an objective but also consistent performance measure across different job-levels and tenure.

This performance measure makes our dataset particularly ideal for this study because we are interested in the change in the *slope* of the dismissal policy. If the unit of the performance measure is not consistent over time, then the change in the slope of the dismissal policy or the wage contract may simply reflect the change in the unit of the performance measure. For example, subjective performance ratings from a supervisor would not be an ideal measure for this type study.

For our analysis, we use either a two-week average performance measure or a 6-month average performance measure[15]. The different time units do not change the qualitative results of the paper

Compensation includes salary, bonus, and overtime payments. The average *6-month* com-

[14] There are four main job-levels. They differ only in the types of claims they process. The tasks are essentially the same cross different job-levels. For more details on the job-levels, see Kwon (1999).

[15] The choice of the time unit is partly because overtime payments are made every two weeks and salaries change, on average, every 6 months.

pensation is about $11,000. Salaries change, on average, every 6 months, either by merit or by promotion. The overtime work hours are measured every two weeks and paid accordingly. The contract is highly personalized. In particular, the dates of the salary change or the dates of the bonus are uniformly distributed throughout the calendar year, and promotions and bonuses are not concentrated at the end of the year or the month.[16]

4.1.3 Turnover

Turnover rates are relatively high. About 32% of the employees in the sample quit during the two and a half year sample period. The total number of employees has been stable during our sample period. In particular, there were no significant layoffs or a bulk hiring in a particular month. Thus, most new hirings are primarily done to fill out the vacancies generated by the dismissals. These features are consistent with the assumptions of our theoretical model. If there were a bulk hiring (e.g. large January hirings), it would imply the economies of scale in hiring and undermine the assumption 1 in our theoretical model.

As Table 3 shows, our data contain detailed information on the nature of the turnovers. In particular, we can distinguish voluntary turnovers from involuntary turnovers. About 65% of the turnovers are voluntary, and the remaining 35% are involuntary. Most involuntary turnovers are performance-related. Since our models are about involuntary turnover, this distinction between involuntary and voluntary turnovers is important. Many previous studies do not distinguish between these (e.g. Chevalier and Ellison (1999)). On the other hand, some of the voluntary turnovers, such as those due to 'more money' or 'job content', can be disguised-involuntary turnovers. Thus, we will analyze both the involuntary turnovers and the total number of turnovers.

[Table 3 here]

[16] For a more detailed description of the wage policy of this firm, see Kwon (1999).

Figure 4 shows the distribution of tenure at the time of terminations. Most turnovers occur in the early part of the employees' tenure regardless of whether they are involuntary. This is consistent with both the sorting model and the incentive model. However, relatively many involuntary turnovers occur later in the agent's career. Assuming that the sorting happens relatively quickly given the high frequency of performance measures, these late involuntary turnovers are likely due to incentive reasons.

[Figure 4 here]

4.1.4 Tenure

Even though the sample period is only two and a half years, the data have the date of hire for each employee. Thus, we can compute tenure. The median tenure of the sample is about 3 years.[17] However, due to a high turnover rate and the short sample period, we do not observe the maximum tenure of those who still remain at the end of our sample period (right-censoring problem). Therefore, the true median tenure is likely to be higher.[18] Indeed, Figure 5 shows that the sample distribution of the tenure is highly skewed. Thus, we should be careful about the selection bias in the following analysis because those who remain after 5 years may have very different characteristics from those who just got hired.

[Figure 5 here]

[17] In Table 1, tenure is measured in units of 2-weeks. Thus, 79 is approximately 3 (≈79/26) years.

[18] A duration estimation controlling for this right-censoring problem predicts that the median duration can be as high as 5 to 7 years, depending on the duration model. The results are not reported in this paper, but are available upon request.

4.2 Productivity and Tenure: Learning-by-Doing

Learning-by-doing is an important feature of our theoretical model. First we study the existence and the extent of learning-by-doing in our data. Figure 6 shows the average performance of the employees is clearly increasing with tenure. However, as noted above, this figure can be also driven by selection bias. Even without learning-by-doing, if unproductive workers quit first, the average productivity of the remaining workers will increase.

[Figure 6 here]

To understand the relationship between productivity and tenure controlling for the individual heterogeneity, we estimate

$$\text{perf}_{it} = a + b(\text{tenure}_{it}) + c(\text{tenure}_{it})^2 + d(\text{controls}_i) \tag{22}$$

where $perf_{it}$ is worker i's performance at time t, $tenure_{it}$ is the tenure measured in 2-week units, and $controls_i$ are other variables that might affect the worker's performance, such as education. In Table 4, we also include the individual worker random effect and fixed effect.

Even after controlling for individual heterogeneity, Table 4 shows significant productivity increase with tenure. The results from the fixed effect estimation suggest that performance can increase for up to 7 years.

[Table 4 here]

4.3 Dismissal and Tenure: Incentive or Sorting?

In Table 5, we analyze how the dismissal probability changes with tenure and performance. The time period is measured in two week periods.[19]. Table 5 estimates the probability of dismissal

[19] We choose a two-week period because we were told that two weeks of unnotified absence can result in a dismissal. Different measures of the time period (e.g. a month or a quarter) do not change the qualitative results

within a two-week interval. Column (1) shows that the overall dismissal probability decreases with the employee's tenure. This is consistent with both the sorting and the incentive models' predictions. Evaluated at the sample median (see Table 2), Column (1) predicts that the probability dismissal (within a two-week time interval) of an agent with one year of tenure is 0.0031. The probability decreases to 0.0023 at five years of tenure and decrease further down to 0.0015 at ten years of tenure. Thus, the dismissal probability decreases by 50% after ten years of tenure.

Column (2) shows that the dismissal probability is also decreasing in the employee's performance. This is also consistent with both the sorting and the incentive models' predictions. However, controlling for performance, tenure is no longer significant, even though the coefficient remains negative. Note that because the average performance increases with tenure, the average dismissal probability will still decrease with tenure.

A key test of the theoretical model is how the slope of the dismissal probability changes with tenure. Column (3) repeats the probit estimation with the interaction term, tenure*performance. The interaction term is negative and significant, suggesting that the dismissal probability becomes *more* sensitive to performance as the agent's tenure increases. Columns (4)-(6) repeat the analysis with all the turnovers, and show that there is no qualitative changes in the results.[20]

[Table 5 here]

These results are consistent with the prediction of our incentive model, but not with the prediction of our sorting model. Of course, this does not mean that 'sorting model' is not relevant for dismissals in this firm.[21] Our results indicate that on average the incentive explanation dominates the sorting explanation in the estimation.

of our analysis.

[20] We also tried the analysis with the performance-related involuntary turnovers only (that is, excluding, e.g., involuntary turnovers due to job-eliminated, and etc.). The qualitative results did not change.

[21] Especially, at low tenure many dismissals are likely to be due to the sorting, while the dismissals at high tenure

Note that because a firm needs strong commitment to use the threat of dismissal as an incentive device, one could be skeptical on the incentive rationale of dismissals. As far as we know, this is the first study that provides a systematic evidence for an incentive rationale of dismissals. As we discussed earlier, most previous studies have not distinguished the incentive rationale and the sorting rationale.

To better illustrate the results, Figure 7(a) shows the *predicted* dismissal probability for a median employee at different levels of tenure using the estimates from Column (3) in Table 5. Figure 7(b) shows the dismissal probability with the mean-adjusted performance, $\tilde{y}_t$. Thus, Figure 7(b) is directly comparable to Figure 3(a). Both 7(a) and 7(b) clearly show that the slope of the dismissal probability becomes steeper as the agent's tenure increases. Furthermore, these figures are close to those in Figure 3, the predictions of the incentive model.

[Figure 7 here]

To measure the economic significance of these estimates, Table 6 shows the change in the dismissal probability when performance decreases from 150 to 100. Approximately 150 is the sample median and 100 marks the first quartile of performance. Even though the level of the dismissal probability decreases with tenure, the change in the probability increases. Especially in percentage terms, the dismissal probability increases by 46% in the first year of tenure, but by 80% in the fifth year and by 108% in the tenth year. The dismissal probability becomes more sensitive to the performance as the tenure increases.

[Table 6 here]

are likely due to the incentive rationale. To test this hypothesis, we have repeated the estimations in Table 5 after splitting the workers into two groups with low tenure and with high tenure. (not reported in the paper) While the coefficient of the interaction term is smaller (in absolute value) for the low-tenure workers, it still remains negative.

4.4 Wage Contract and Tenure

The incentive model predicts that the average wage will increase with tenure even if there is no long-term contract and as well human capital is firm-specific. The slope of the wage contract will increase with tenure as well. Table 7 estimates the wage function. Columns (1) and (2) show that the wage is increasing with tenure. Column (3) shows that for a given level of tenure, the wage is also increasing with performance. Column (4) adds the interaction term and shows that the *slope* of the wage function is also increasing with tenure, which is consistent with the predictions of our theoretical model.

[Table 7 here]

In our theoretical model, the slope of the wage function increases with tenure because the principal relies more heavily on monetary incentives as the threat of dismissal (i.e. replacement of existing workers) becomes more expensive due to learning-by-doing. However, there are many other factors that can contribute to the increase in the slope of the wage function, such as career concerns, or long-term contracts. Decomposing these effects is certainly an interesting topic, but beyond the scope of this paper.

4.5 Alternative Explanations

Previous sections show that the changes in the contract with employees' tenure, especially the change in the dismissal probability, are consistent with the predictions of the incentive model. However, it is possible that even in a sorting model, the slope of the dismissal probability increases with tenure (in absolute value). If the variance of the error term (σ_ϵ^2) decreases with tenure fast enough, then from (7), the slope of the dismissal probability in the sorting model ($= -\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_\epsilon^2+\sigma_t^2}$) can increase with tenure (in absolute value). To test this hypothesis, Figure 8 shows that the standard deviation of performance for every six months of tenure. Since this standard deviation includes the variation of individual ability, Figure 8 also shows the standard deviation of the residual performance after controlling for individual fixed effects and tenure. (from Table 4 (5)). The figure shows that the variance is actually slightly increasing with tenure. Thus, under the sorting model, the slope of the dismissal probability should decrease (in absolute value) with tenure even faster than the case with a constant σ_{ϵ}^{2}.

[Figure 8 here]

Another possibility is that the interaction term (tenure*perf) in Table 5 may act as a tenure-squared term because tenure and performance are highly correlated from learning-by-doing. Also since performance tends to increase with tenure, our results may be picking up some type of shift that occurs because what is considered good performance varies with tenure due to learning-by-doing. One way for controlling this in the regression is to substitute for raw performance the difference between raw performance and tenure-adjusted average performance. Therefore, in Table 8, we repeat the probit estimation with a tenure-squared term and the difference measure of performance.

[Table 8 here]

Column (1) shows that even after controlling for the tenure-squared, the interaction term remains negative and significant. Also, column (2) repeats Column (1) with measuring the performance by the difference between the performance and the average performance of the workers with the same tenure. It shows that changing the performance measure to the difference does not affect the results.

One might be also concerned that the unobserved heterogeneity of the employees is causing some bias in our estimates. Thus, we also control the employee random effect in column (3). The results are still robust, however.

Since these workers are promoted to higher job-level as tenure increases, one can also imagine that the tenure variable may capture the different nature of the claims processed in different job-levels, even though 'the weighting system' on the performance intends to take care of this problem. Column (4) in Table 8 adds the job levels and their interactions with the performance. However, none of the job-level variables are significant, and the interaction term between performance and tenure still remains negative and significant.

5 Conclusion

This paper shows that dismissals due to poor performance can arise from two reasons: 'incentives' and 'sorting'. These are two very different reasons for dismissals both practically and theoretically. However, there has been few studies that try to distinguish them. While there are many studies that show the negative relationship between the dismissal probability and performance, this finding is consistent with both the sorting model and the incentive model. We show that in a dynamic model with learning-by-doing, one can distinguish these two alternative explanations for dismissals by looking at how the sensitivity (or slope) of the dismissal probability with respect to performance changes over time. Using detailed personnel records, we find the support for the incentive model.

The empirical support for the incentive model is also important because a typical agency model would predict that the threat of dismissal is not an optimal incentive device if the principal can use performance-based wage contract. Furthermore, the dismissal requires a large commitment of the firm since it may have to fire a highly experienced worker. Thus, it is not obvious whether firms would actually use the threat of dismissal as an incentive device. This paper shows that the threat of dismissal is an optimal incentive device in a repeated moral hazard model with limited liability even when firms can lower the wages based on performance. Our empirical evidence supports this result.

The empirical analysis also provides some evidence for the sorting model as well. However, it does not tell us how many of the dismissals are due to 'sorting' and how many are due to 'incentives'. To answer this question, we need a model of dismissals that integrates both the incentive and the sorting model, then will have to rely on a structural estimation. We are pursuing this in a future research.

References

[1] Abbring, J.H. et. al., 2003, "Moral Hazard and Dynamic Insurance Data," forthcoming, *Journal of the European Economic Association*

[2] Calvo, G.A., 1985, "The Inefficiency of Unemployment: The Supervision Perspective," *Quarterly Journal of Economics*, 100, 379-387

[3] Chevalier, J. and G. Ellison, 1999, "Career Concerns of Mutual Fund Managers," *Quarterly Journal of Economics*, 389-432

[4] Conyon, M.J., 1998, "Directors' Pay and Turnover: An Application to A Sample of Large UK Firms," *Oxford Bulletin of Economics and Statistics*, 60, 4, 485-507

[5] Denis, D. J. and D. K. Denis, 1995, "Performance Changes Following Top Management Dismissal," *Journal of Finance*, 50, 1029-1057

[6] Jensen, M.C. and K.J. Murphy, 1990, "Performance Pay and Top Management Incentives," *Journal of Political Economy*, 98, 225-64

[7] Jovanovic, B. and Y. Nyarko, 1996, "Learning by Doing and The Choice of Technology," *Econometrica*, 64, 1299-1310

[8] Kaplan, S., 1994, "Top executive rewards and firm performance: A comparison of Japan and the United States," *Journal of Political Economy*, 102, 510-546

[9] Kuhn, P., 1986, "Wages, Effort, and Incentive Compatibility in Life-Cycle Employment Contracts," *Journal of Labor Economics*, 4, 28-49

[10] Kwon, I., 1999, "Wages, Incentives, and Promotions: Theory and Evidence," mimeo

[11] _______, 2003, "Threat of Dismissal and Authority," mimeo

[12] Levin, J., 2000, "Relational Incentive Contracts," forthcoming, *American Economic Review*

[13] MacLeod, W.B., 2001, "On Optimal Contracting with Subjective Evaluation," forthcoming, *American Economic Review*

[14] MacLeod, W.B. and J.M. Malcomson, 1989, "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment," *Econometrica*, 57, 447-480

[15] Mori, P.A., 1998, "Task Shifts vs. Termination as Devices for Eliciting Optimal Effort Supply," *Journal of Economics and Management*, 7, 33-65

[16] Oi, W. Y. and T.L. Idson, 1999, "Firm Size and Wages," *Handbook of Labor Economics,* Vol 3, Elsevier Science, chapter 33, pp. 2166-2114.

[17] Prendergast, C., 2000, "What Trade-Off of Risk and Incentives?," *American Economic Review*, 90, 421-25

[18] Rogerson, W., 1985, "Repeated Moral Hazard," *Econometrica*, 53, 69-76

[19] Shapiro, C. and J. Stiglitz, 1984, "Worker Unemployment as a Worker Discipline Device," *American Economic Review*, 74, 433-444

[20] Sparks, R., 1986, "A Model of Involuntary Unemployment and Wage Rigidity: Worker Incentives and the Threat of Dismissal," *Journal of Labor Economics*, 4, 560-581

[21] Weisbach, M., 1988, "Outside directors and CEO turnover," *Journal of Financial Economics*, 20, 431-460

Appendix A Sorting Model with General Ability

Suppose that $\theta = \theta_G + \theta_F$ where θ_G is the general ability and θ_F is the firm-specific matching quality. Let's assume that $\theta_F = p\theta$ and $\theta_G = (1-p)\theta$ where $0 < p \leq 1$. Thus, p is the market's belief on the proportion of firm-specific matching quality in θ. Similarly, suppose that $l = l_G + l_F$ where l_G is the general human capital accumulation and l_F is the firm-specific human capital accumulation. Let's assume that $l_F = rl$ and $l_G = (1-r)l$. Furthermore, assume that $r > p$[22]. Thus, as the tenure of an employee increases, the firm-specific human capital takes larger portion of an employee's total human capital. Then, given y_t and m_t,

$$\begin{aligned} w_{t+1} &= E(\theta_G|y_t, m_t) + tl_G = (1-p)E(\theta|y_t, m_t) + tl_G \\ &= (1-p)\left[\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} - tl\right] + t(1-r)l \\ &= (1-p)\left[\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2}\right] - tl(r-p) \end{aligned}$$

Then, the principal's maximization problem is now as follows:

$$\begin{aligned} \max_{0 \leq q_t(y_t;m_t)) \leq 1} E[\pi_{t+1}|y_t] &= q_t(y_t;m_t)0 + (1-q_t(y_t;m_t))\left(\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l - w_{t+1}\right) - \frac{\phi N}{2} q_t(y_t;m_t)^2 \\ &= q_t(y_t;m_t)0 + (1-q_t(y_t;m_t))\left(p\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l + tl(r-p)\right) - \frac{\phi N}{2} q_t(y_t;m_t)^2 \end{aligned}$$

From the FOC, it is straightforward to show that the optimal dismissal policy is:

$$q_t^*(y_t;m_t) = \begin{cases} 1 & \text{if } y_t < y_t^1(m_t) \\ -\frac{1}{\phi N}(p\frac{\sigma_\epsilon^2 m_t + \sigma_t^2 y_t}{\sigma_\epsilon^2 + \sigma_t^2} + l + tl(r-p)) & \text{if } y_t^1(m_t) \leq y_t \leq y_t^0(m_t) \\ 0 & \text{if } y_t > y_t^0(m_t) \end{cases}$$

where y_t^1 and y_t^0 are similarly defined as before. Then, from the proof of Proposition1, it is straightforward to see that the Proposition 1 continues to hold.

Appendix B

[22] Even if $p > r$, the qualitative results on the slope of the dismissal probability do not change. That is, Proposition 1(i) continues to hold. However, if $p - r$ is positive and large enough, the level of the dismissal probability may increase with tenure.

Proof of Proposition 1 (i) From (7), for $0 < q_t^*(y;m) < 1$, we have $\frac{\partial q_t^*(y;m)}{\partial y} = -\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_t^2+\sigma_\epsilon^2} < 0$. Since $\sigma_{t+1}^2 = \frac{\sigma_t^2\sigma_\epsilon^2}{\sigma_t^2+\sigma_\epsilon^2} < \sigma_t^2$ for all $t = 1, 2, 3, ...$, $\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_t^2+\sigma_\epsilon^2}$ decreases in t. Therefore, $\frac{\partial^2 q_t^*(y;m)}{\partial t \partial y} = -d(\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_t^2+\sigma_\epsilon^2})/dt > 0$.

(ii) First, consider the change of the dismissal probability for a given y. Holding m constant, only σ_t^2 changes with t. From (7),

$$\begin{aligned}\frac{\partial q_t(y;m)}{\partial t} &= -\frac{1}{\phi N}\left[\frac{(\sigma_\epsilon^2+\sigma_t^2)y-(\sigma_\epsilon^2 m+\sigma_t^2 y)}{(\sigma_\epsilon^2+\sigma_t^2)^2}\frac{\partial \sigma_t^2}{\partial t}\right] \\ &= -\frac{1}{\phi N}\left[\frac{\sigma_\epsilon^2(y-m)}{(\sigma_\epsilon^2+\sigma_t^2)^2}\frac{\partial \sigma_t^2}{\partial t}\right]\end{aligned}$$

From above, $\frac{\partial \sigma_t^2}{\partial t} < 0$. Also, from (7), $y < 0$ for $0 < q_t^*(y;m) < 1$. Therefore, if $m > 0$, then $\frac{\partial q_t(y;m)}{\partial t} < 0$.

Now consider the change of the expected dismissal probability. Holding m constant, only σ_t^2 changes with t. Denote the normal probability distribution function with mean m and the variance σ^2 by $F(y|m,\sigma^2)$. Then,

$$\begin{aligned}E_{y_t}\left[q_t(y_t;m)\right] &= \int_{-\infty}^{y_t^1} q_t(y_t;m)dF(y_t|m,\sigma_t^2) + \int_{y_t^1}^{y_t^0} q_t(y_t;m)dF(y_t|m,\sigma_t^2) \\ &= F(y_t^1|m,\sigma_t^2) + q_t(y_t;m)F(y_t|m,\sigma_t^2)\Big|_{y_t^1}^{y_t^0} - \int_{y_t^1}^{y_t^0} F(y_t|m,\sigma_t^2)\frac{\partial q_t(y_t;m)}{\partial y_t}dy_t \\ &= -\int_{y_t^1}^{y_t^0} F(y_t|m,\sigma_t^2)\frac{\partial q_t(y_t;m)}{\partial y_t}dy_t \\ &= \frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_t^2+\sigma_\epsilon^2}\int_{y_t^1}^{y_t^0} F(y_t|m,\sigma_t^2)dy_t\end{aligned}$$

Note that $F(y_t|m,\sigma_t^2) > 0$ and $\frac{\partial F(y_t|m,\sigma^2)}{\partial \sigma^2} > 0$ for all $y_t < m$. As σ_t^2 increases, $\frac{1}{\phi N}\frac{\sigma_t^2}{\sigma_t^2+\sigma_\epsilon^2}$ increases Therefore, $\frac{\partial E_{y_t}[q_t(y_t;m)]}{\partial \sigma_t^2} > 0$. Since σ_t^2 decreases with t, $E_{y_t}\left[q_t(y_t;m)\right]$ decreases with t.■

Proof of Lemma 1 First, note that if $v_2(y_2) = 0$ for all y_2, then the incentive constraint (10) is violated. Therefore, $v_2(y_2) > 0$ at least for some y_2. Then, from (10) and (11), $\int v_2(y_2)f(y_2;\theta_2,a_H)dy_2 - H > 0$.■

Proof of Lemma 2 Define $\tilde{y}_2 = y_2 - \theta_2$. Then, from (8), for any θ_2

$$\begin{aligned}V_2^* &= \int v_2(y_2;\theta_2)f(y_2;\theta_2,a_H)dy_2 - H \\ &= \int v_2(\tilde{y}_2;0)f(\tilde{y}_2;0,a_H)dy_2 - H\end{aligned}$$

Therefore, V_2^* does not depend on θ_2.■

Proof of Proposition 2 First, like lemma 1, the participation constraint is not binding from (16) and (17). It is easy to see that the incentive constraint is binding. Then, for any given y_1, assuming an interior solution, the FOC for $v_1(y_1)$ is as follows:

$$-h'(v_1(y_1))f(y_1|\theta_1, a_H) + \lambda_1[f(y_1|\theta_1, a_H) - f(y_1|\theta_1, a_L)] = 0$$

where λ_1 is the Lagrange multiplier for the incentive constraint (16).

Therefore,

$$v_1^*(y_1;\theta_1) = \begin{cases} g(\lambda_1(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)})) & \text{if } y \geq \overline{y}_1(\theta_1) \\ 0 & \text{if } y < \overline{y}_1(\theta_1) \end{cases}$$

where $g \equiv h'^{-1}$ and $\bar{y}_1(\theta_1)$ satisfies $g(\lambda_1(1 - \frac{f(\bar{y}_1|\theta_1,a_L)}{f(\bar{y}_1|\theta_1,a_H)})) = 0$.

We now show that $\bar{y}_1(\theta_1)$ exists and unique. First, suppose that $\bar{y}_1(\theta_1)$ does not exist. Then, since $g(\lambda_1(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)}))$ is continuous and strictly increasing, $g(\lambda_1(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)})) > 0$ for all y_1. Thus, the limited liability constraint is not binding. However, if the limited liability constraint is not binding, then the participation constraint must be binding. A contradiction.

Since $g(\lambda_1(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)}))$ is strictly increasing, $\bar{y}_1(\theta_1)$ is also unique. Substituting this into the incentive constraint determines λ_1 as a function of θ_1.

For the following proof, define $\hat{y}$ such that $f(\hat{y}|\theta_1, a_L) = f(\hat{y}|\theta_1, a_L)$. Note that from the FOC, $\lambda_1(1 - \frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)}) > 0$ since $h' > 0$. That is, the optimal contract is defined for $y > \hat{y}$. In particular, we must have $\bar{y}_1 \geq \hat{y}$.

For any given y_1, assuming the interior solution, the FOC for $q_1(y_1)$ is as follows:

$$[\delta\Pi_2^*(k_0) - \delta\Pi_2^*(k_{\tau+1}) - \phi N q_1(y_1)]f(y_1|\theta_1, a_H) + \lambda_1\delta V_2^*[-f(y_1|\theta_1, a_H) + f(y_1|\theta_1, a_L)] = 0$$

Therefore,

$$q_1^*(y_1;\theta_1) = \begin{cases} 1 & \text{if } y_1 \leq y^1(\theta_1) \\ \frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)} - 1) - \Delta_{\tau+1}(\theta_1)] & \text{if } y^1(\theta_1) < y_1 \leq y^0(\theta_1) \\ 0 & \text{if } y_1 > y^0(\theta_1) \end{cases}$$

where $\Delta_{\tau+1}(\theta_1) \equiv \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$. $y^1(\theta_1)$ and $y^0(\theta_1)$ are defined such that $\frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y^1|\theta_1,a_L)}{f(y^1|\theta_1,a_H)} - 1) - \Delta_{\tau+1}(\theta_1)] = 1$ and $\frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y^0|\theta_1,a_L)}{f(y^0|\theta_1,a_H)} - 1) - \Delta_{\tau+1}(\theta_1)] = 0$. Since $\frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)} -$

$1) - \Delta_{\tau+1}(\theta_1)]$ is strictly decreasing in y, from assumption 2, y^1 and y^0 are well-defined and unique. Also $y^0 > y^1$.

At $y_1 = \hat{y}$, $\frac{\delta}{\phi N}[\lambda_1(\theta_1)V_2^*(\frac{f(y_1|\theta_1,a_L)}{f(y_1|\theta_1,a_H)} - 1) - \Delta_{\tau+1}(\theta_1)] < 0$. Therefore, $y_0 < \hat{y}$. Since $\bar{y}_1 \geq \hat{y}$, we have $\bar{y}_1 > y^0 > y^1$.■

Proof of Proposition 3 If we rewrite the optimal contract in terms of $\tilde{y}_1 = y_1 - \theta_1$, then from (8), $v_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}) = g(\lambda_1(1 - \frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)}))$ and $q_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}) = \frac{\delta}{\phi N}[\lambda_1 V_2^*(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} - 1) - \Delta_{\tau+1}]$. Note that since V_2^* does not depend on $\theta_2 (= k_{\tau+1})$ from lemma 2, we can treat V_2^* as a constant. As τ increases, from lemma 3, $\Delta_{\tau+1} = \Pi_2^*(k_{\tau+1}) - \Pi_2^*(k_0)$ increases. Then, from the binding incentive constraint, λ_1 also changes. To prove (i) and (ii), it is sufficient to show that $\frac{\partial \lambda_1}{\partial \Delta_{\tau+1}} > 0$. Then, from (19) and (20), the slopes of both the dismissal probability and the wage function become steeper as τ increases.

For simplicity, denote $\bar{y}_1(0), y^0(0)$, and $y^1(0)$ by $\bar{y}_1$, y^0, and y^1 without the argument. Since $\bar{y}_1 > y^0 > y^1$ from Proposition 2, we can rewrite the incentive constraint (16) as follows:

$$
\begin{aligned}
IC(\lambda_1, \Delta_{\tau+1}) \;&\equiv\; \int [v_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}) + (1 - q_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}))V_2^*]\,(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\tilde{y}_1 - H \\
&=\; \int_{y^1}^{y^0} [(1 - q_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}))V_2^*](f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\tilde{y}_1 + \int_{y^0}^{y^1} V_2^*(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 \\
&\quad + \int_{\bar{y}}^{\infty} [v_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1}) + V_2^*]\,(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 - H \\
&=\; 0
\end{aligned}
$$

Claim 1 $\frac{\partial IC(\lambda_1, \Delta_{\tau+1})}{\partial \lambda_1} > 0$.

$$
\begin{aligned}
\frac{\partial IC(\lambda_1, \Delta_{\tau+1})}{\partial \lambda_1} \;&=\; -\int_{y^1}^{y^0} \frac{\partial q_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1})}{\partial \lambda_1} V_2^*(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 \\
&\quad + \int_{\bar{y}}^{\infty} \frac{\partial v_1^*(\tilde{y}_1; \lambda_1, \Delta_{\tau+1})}{\partial \lambda_1}(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 \\
&=\; -\int_{y^1}^{y^0} \frac{\delta}{\phi N}(\frac{f(\tilde{y}_1|0, a_L)}{f(\tilde{y}_1|0, a_H)} - 1)(V_2^*)^2(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 \\
&\quad + \int_{\bar{y}}^{\infty} g'(\lambda_1(1 - \frac{f(\tilde{y}_1|0, a_L)}{f(\tilde{y}_1|0, a_H)}))(1 - \frac{f(\tilde{y}_1|0, a_L)}{f(\tilde{y}_1|0, a_H)})(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))d\widetilde{y}_1 \\
&>\; 0
\end{aligned}
$$

The inequality is due to $g' > 0$ and $(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} - 1)(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L)) = -\frac{1}{f(\tilde{y}_1|0,a_H)}(f(\tilde{y}_1|0, a_H) - f(\tilde{y}_1|0, a_L))^2 < 0$. Note that the changes of $\bar{y}_1$, y^0, and y^1 all cancel out.

Claim 2 $\frac{\partial IC(\lambda_1,\Delta_{\tau+1})}{\partial\Delta_{\tau+1}} < 0$.

$$\begin{aligned}\frac{\partial IC(\lambda_1,\Delta_{\tau+1})}{\partial\Delta_{\tau+1}} &= -\int_{y^1}^{y^0}\frac{\partial q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})}{\partial\Delta_{\tau+1}}V_2^*(f(\tilde{y}_1|0,a_H)-f(\tilde{y}_1|0,a_L))d\widetilde{y}_1 \\ &= \int_{y^1}^{y^0}\frac{\delta}{\phi N}V_2^*(f(\tilde{y}_1|0,a_H)-f(\tilde{y}_1|0,a_L))d\widetilde{y}_1 < 0\end{aligned}$$

The last inequality is due to $f(\tilde{y}_1|0,a_H)-f(\tilde{y}_1|0,a_L)<0$ for $y^1<y<y^0(<\hat{y})$. Recall that $\hat{y}$ is defined such that $f(\hat{y}|\theta_1,a_L)=f(\hat{y}|\theta_1,a_L)$.

Claim 3 $\frac{\partial\lambda_1}{\partial\Delta_{\tau+1}} > 0$.

From the implicit function theorem,

$$\frac{d\lambda_1}{d\Delta_{\tau+1}} = -(\frac{\partial IC}{\partial\Delta_{\tau+1}})/(\frac{\partial IC}{\partial\lambda_1}) > 0$$

Therefore, as τ increases, λ_1 increases as well. Then, from (19) and (20), the slopes of both the dismissal probability and the wage function increase (in absolute value). That is, $\frac{\partial^2 q_1(\tilde{y}_1;\tau)}{\partial\tau\partial\tilde{y}_1}<0$ and $\frac{\partial^2 v_1(\tilde{y}_1;\tau)}{\partial\tau\partial\tilde{y}_1}>0$.

Claim 4 $\frac{\partial v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})}{\partial\tau}>0$ and $\frac{\partial E[v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau}>0$.

Since $(1-\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)})$ does not change with τ, if λ_1 increases, $\bar{y}$ has to decrease. Since the slope is also larger for any $\tilde{y}_1$, $v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})$ must increase with τ. Furthermore, because $v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})$ increases with τ for any given $\tilde{y}_1$, $\frac{\partial E[v_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau}>0$.

Claim 5 $\frac{\partial E[q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau}<0$.

From (20), the change in τ affects the dismissal probability through λ_1 and $\Delta_{\tau+1}$. Therefore,

$$\frac{\partial E[q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau} = \frac{\delta}{\phi N}\left[\int_{y^1}^{y^0}(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)}-1)V_2^*f(\tilde{y}_1|0,a_H)d\widetilde{y}_1\right]\frac{d\lambda_1}{d\tau}-\frac{\delta}{\phi N}\left[\int_{y^1}^{y^0}f(\tilde{y}_1|0,a_H)d\widetilde{y}_1\right]\frac{d\Delta_{\tau+1}}{d\tau}$$

Suppose that λ_1 in the wage function is fixed. Then, from Claim 3,

$$\frac{d\lambda_1}{d\Delta_{\tau+1}} = -\frac{\int_{y^1}^{y^0}(f(\tilde{y}_1|0,a_H)-f(\tilde{y}_1|0,a_L))d\widetilde{y}_1}{-\int_{y^1}^{y^0}(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)}-1)V_2^*(f(\tilde{y}_1|0,a_H)-f(\tilde{y}_1|0,a_L))d\widetilde{y}_1} \tag{A.1}$$

For the simplicity of exposition, denote $A=\int_{y^1}^{y^0}(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)}-1)V_2^*f(\tilde{y}_1|0,a_H)d\widetilde{y}_1$, $B=\int_{y^1}^{y^0}(\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)}-1)V_2^*f(\tilde{y}_1|0,a_L)d\widetilde{y}_1$, $C=\int_{y^1}^{y^0}f(\tilde{y}_1|0,a_H)d\widetilde{y}_1$, and $D=\int_{y^1}^{y^0}f(\tilde{y}_1|0,a_L)d\widetilde{y}_1$. Note that $A<B$ and $C<D$

since $f(\tilde{y}_1|0,a_H) < f(\tilde{y}_1|0,a_L)$ for $y^1 < \tilde{y}_1 < y^0$. Then, $\frac{d\lambda_1}{d\Delta_{\tau+1}} = \frac{D-C}{B-A}$. Therefore, we can rewrite as follows:

$$\begin{aligned}\frac{\partial E[q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau} &= \frac{\delta}{\phi N}\left[A * \frac{d\lambda_1}{d\tau} - C * \frac{d\Delta_{\tau+1}}{d\tau}\right] \\ &= \frac{\delta}{\phi N}\left[A\frac{D-C}{B-A} - C\right]\frac{d\Delta_{\tau+1}}{d\tau}\end{aligned}$$

Also define $A' \equiv \int_{y^1}^{y^0} \frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} f(\tilde{y}_1|0,a_H)d\widetilde{y}_1$ and $B' \equiv \int_{y^1}^{y^0} \frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} f(\tilde{y}_1|0,a_L)d\widetilde{y}_1$. Then, the following holds:

$$\begin{aligned}A\frac{D-C}{B-A} - C &< 0 \Longleftrightarrow AD - BC = A'D - B'C < 0 \\ &\Longleftrightarrow \frac{A'}{C} - \frac{B'}{D} < 0 \\ &\Longleftrightarrow \frac{\int_{y^1}^{y^0} \frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} f(\tilde{y}_1|0,a_H)d\widetilde{y}_1}{\int_{y^1}^{y^0} f(\tilde{y}_1|0,a_H)d\widetilde{y}_1} < \frac{\int_{y^1}^{y^0} \frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} f(\tilde{y}_1|0,a_L)d\widetilde{y}_1}{\int_{y^1}^{y^0} f(\tilde{y}_1|0,a_L)d\widetilde{y}_1}\end{aligned}$$

The last inequality holds because $\frac{f(\tilde{y}_1|0,a_L)}{f(\tilde{y}_1|0,a_H)} > 1$ for $y^1 < \tilde{y}_1 < y^0$ and strictly decreasing. Since $\frac{d\Delta_{\tau+1}}{d\tau} > 0$ from lemma 3, $\frac{\partial E[q_1^*(\tilde{y}_1;\lambda_1,\Delta_{\tau+1})]}{\partial\tau} < 0$ holding λ_1 in the wage function constant. If we allow λ_1 in the wage contract to change, then λ_1 will not increase as much as (A.1). Thus, the expected dismissal probability will decrease even further.■

Proof of Lemma 4 Suppose that $V_t^* \geq V_{t+1}^*$ for any given t. Since the optimization problems in each period is identical except for Δ_t (which is increasing in t), $V_t^* \geq V_{t+1}^*$ for all t. For a moment, assume that Δ_t is constant and that only V_t^* is decreasing in t. From (20) both the level and the slope of the the dismissal probability decrease with t. Therefore, in order to satisfy the incentive constraint, λ_t has to increase with t. Therefore, both the slope and the level of the wage function should increase. Now if we allow Δ_t to increase in time, from proposition 3, it also increases the wages and decreases the expected dismissal probability. Then, however, the continuation payoff of the agent must increase with t. A contradiction.■

Proof of Proposition 4 As t increases, both Δ_t and V_t^* increase. In proposition 3, we have already showed that for a given V_t^*, the increase in Δ_t raises λ_t. Therefore, from (20), it is enough to show that for given Δ_t, the increase in V_t^* raises $V_t^*\lambda_t$. With a slight abuse of notation, define $s \equiv V_t^*\lambda_t$. Then, $\lambda_t = \frac{s}{V_t^*}$. It is straightforward to show that $\frac{\partial IC(s,V_t^*)}{\partial s} > 0$ and $\frac{\partial IC(s,V_t^*)}{\partial V_t^*} < 0$. Therefore, $\frac{ds}{dV_t^*} > 0$.■

TABLE 1 Predictions on Dismissal Probability

	Sorting Model	Incentive Model
$\frac{\partial q(y,t)}{\partial y}$	$-$	$-$
$\frac{\partial q(y,t)}{\partial t}$	$-$	?
$\frac{\partial E_y[q(y,t)]}{\partial t}$	$-$	$-$
$\frac{\partial^2 q(y,t)}{\partial t \partial y}$	$+$	$-$

Note: $q(y,t)$ = dismissal probability for performance y at tenure t.

TABLE 2 Description of Variables and Summary Statistics

Variable	Description	Obs	Mean	Std. Dev.	5%	50%	95%
quit	=1 if worker is dismissed	3,231	0.32	0.47	0	0	1
educ	education (year)	3,045	12.92	1.57	12	12	16
age	age (year)	3,231	31.07	7.36	22	30	45
female	=1 if female	3,231	0.91	0.29	0	1	1
marital	=1 if married	3,209	0.56	0.50	0	1	1
wage[a]	6-month total wage	10,522	10,995.87	2,797.37	7,850	10,287.54	16,246.77
perf[b]	performance	112,071	174.01	108.04	23.52	159.19	359.52
tenure[c]	tenure (2-weeks)	112,071	118.88[d]	118.76	4	79	365

[a]: The wage is the six month sum of salary, bonus, and overtime payments.

[b]: Performance is the two-week average of the weighted number of claims processed a day.

[c] : Tenure is measured in units of two weeks. (Thus, 79 is approximately 3 ($\approx$ 79/26) years.)

[d] : Due to right-censoring, the tenure average is underestimated.

TABLE 3 Turnover Reasons

Turnover Reason	Frequency	Percent
VOL: ENTERING NEW FIELD	132	12.8
VOL: FAMILY OBLIGATIONS	86	8.4
VOL: SPOUSE RELOCATED	65	6.3
VOL: MORE MONEY	58	5.3
VOL: ADVANCEMENT OPPORTUNITY	55	5.3
VOL: JOB CONTENT	49	4.8
VOL: RETURNING TO SCHOOL	38	3.7
VOL: JOB ABANDONMENT	36	3.5
VOL: LOCATION	22	2.1
VOL: HEALTH	11	1.1
VOL: FAILED TO RETURN	10	1.0
VOL: WORKLOAD	6	0.6
VOL: JOB CHALLENGE	7	0.7
VOL: COMMUTING DIFFICULTIES	5	0.5
VOL: CONFLICT W/ SUPERVISOR	3	0.3
VOL: BENEFITS	1	0.1
VOL: WORKING CONDITIONS	3	0.3
VOL: OTHER	87	8.4
VOL: TOTAL	674	65.5
INVOL: PERFORMANCE	162	15.7
INVOL: PUT ON PROBATION	45	4.4
INVOL: JOB ELIMINATED	41	4.0
INVOL: ATTENDANCE	34	3.5
INVOL: UNETHICAL CONDUCT	8	0.8
INVOL: VIOLATION OF PUBLISHED RULES	4	0.4
INVOL: FRAUD OR DISHONESTY	3	0.3
INVOL: OTHERS.	59	5.7
INVOL: TOTAL	355	34.5
TOTAL	1,029	100

TABLE 4 Productivity and Tenure

(dependent variable: perf_{it})

	OLS	Random-Effect	Fixed-Effect
	(1)	(2)	(3)
tenure	.632455***	.813629***	.7655434***
	(.0079041)	(.0180588)	(.0178769)
tenure^2	-.0009719***	-.0016582***	-.0019299***
	(.0000182)	(.0000416)	(.0000465)
age	-.8144315***	-1.07978***	
	(.0513373)	(.215501)	
edu	-.1208695	.4105862	
	(.2099944)	(.9645939)	
female	9.992435***	10.67372**	
	(1.210798)	(5.295515)	
marital	-.1512363	5.087046*	
	(.6658425)	(3.07123)	
#(obs)	103,645	103,645	103,645

* : significant at 10%, ** : significant at 5%, *** : significant at 1%.

[a] : $\text{vol:term}_i = 1$ if a worker i quits voluntarily, =0 otherwise.

[b] : $\text{invol:term}_i = 1$ if a worker i quits involuntarily, =0 otherwise.

TABLE 5 Turnover and Tenure[a]

(dependent variable: $T_{it} = 1$ if worker i gets dismissed at time t, $= 0$ otherwise)

	Involuntary Turnover			All Turnover		
	(1)	(2)	(3)	(4)	(5)	(6)
tenure	-.0009925***	-.0003069	.0007529**	-.0021633***	-.001508***	-.0000141
	(.0002216)	(.0002072)	(.0003323)	(.000186)	(.0001753)	(.0002903)
perf		-.0032223***	-.0022846***		-.0024799***	-.0015493***
		(.0002723)	(.0003718)		(.0001953)	(.0002619)
tenure*perf			-.00000925***			-.0000122***
			(.00000234)			(.00000208)
educ	-.046976***	-.0467884***	-.0474445***	-.0009547	.0001685	-.0005348
	(.0134497)	(.0138367)	(.01391)	(.0079861)	(.0080505)	(.0081483)
age	.0098537***	.0082021***	.0085876***	.0051206***	.0032756*	.0036811*
	(.0029201)	(.0028886)	(.0029141)	(.0019823)	(.0019481)	(.0019793)
constant	-1.965281***	-1.55438***	-1.643303***	-2.143059***	-1.820695***	-1.915419***
	(.2058432)	(.2124916)	(.2152593)	(.1283227)	(.1304188)	(.133939)

* : significant at 10%, ** : significant at 5%, *** : significant at 1%.

[a] : All regressions also include time, female, and marital dummy variables.

[b] : Tenure is measured in units of two-weeks. Twenty six roughly corresponds to one year.

[c] : perf. is measured by the two-week average.

[d] : Standard errors are in parentheses. They are adjusted for clustering.

Table 6 (Involuntary) Dismissal Probability and Tenure

	Dismissal Probability		
Tenure	Perf=150	Perf=100	Change (%)
1 year	0.0026	0.0038	0.0012 (46%)
5 year	0.0021	0.0038	0.0017 (80%)
10 year	0.0016	0.0034	0.0018 (108%)

TABLE 7 Wage Contract and Tenure[a]

(dependent variable=wage[b])

	(1)	(2)	(3)	(4)
tenure(6mo.)	235.4605***	464.5811***	423.6217***	411.0472***
	(6.7745)	(11.7592)	(12.1055)	(14.8554)
$tenure^2$		-7.8233***	-6.9491***	-7.2557***
		(.4318)	(.4250)	(.41919)
perf.			5.0738***	3.7557***
			(.3967)	(.4329)
tenure*perf				.1223**
				(.0521)
educ	109.2255***	116.1456***	116.822***	116.4733***
	(25.0611)	(22.0485)	(21.3054)	(21.4309)
age	.8515	-3.9303	-.5901	-1.1889
	(5.6079)	(5.1999)	(4.8102)	(4.8073)
R^2	0.5841	0.6551	0.6834	0.6851
#(obs)	9925	9925	9925	9925

* : significant at 10%, ** : significant at 5%, *** : significant at 1%.

[a] : The regression runs include time, female, and marital dummy variables which are not reported in the table above.

[b] : The wage is measured by the sum of 6-month salary, bonus, and overtime payment.

[c] : The tenure is measured in units of six month. Thus, ten is approximately 5 years.

[d] : Standard errors are in parentheses. They are adjusted for clustering.

TABLE 8 Probit Estimation with Random Effect

(dependent variable: $T_{it} = 1$ if worker i is terminated involuntarily at time t, $= 0$ otherwise)[a]

	(1)	(2)	(3)	(4)
tenure[b]	.00101	-.003006***	.0005199	.0027288**
	(.00079)	(.000492)	(.0005558)	(.001158)
tenure^2	2.69e-07	3.88e-06***	6.32e-07	-3.15e-06
	(1.81e-06)	(1.10e-06)	(1.19e-06)	(2.54e-06)
perf.[c]	-.00225***		-.0022533***	-.0018931***
	(.00041)		(.0003749)	(.0004521)
tenure*perf.	-.0000109***		-9.28e-06***	-6.60e-06*
	(2.61e-06)		(2.62e-06)	(3.45e-06)
diff.[d]		-.00302***		
		(.000409)		
tenure*diff		-4.60e-06*		
		(2.52e-06)		
educ	-.04312***	-.04641***	-.0474185***	-.0429463***
	(.01419)	(.01334)	(.0133205)	(.014279)
age	.0084***	.009***	.0086418***	.0079535***
	(.0029)	(.0028)	(.0028126)	(.0029637)
constant	-1.7011***	-1.9552***	-1.640344***	-1.723224***
	(.2129)	(.20128)	(.2043405)	(.2144275)
Job-Levels[e]	no	no	no	yes
Random Effect	no	no	yes	yes

* : significant at 10%, ** : significant at 5%, *** : significant at 1%.

[a] : All regressions also include time, female, and marital dummy variables. Standard errors are in parentheses. They are adjusted for clustering.

[b] : 'tenure' is measured by two-weeks. Thus, roughly 26 corresponds to one year.

[c] : 'perf.' is measured by the two-weeks average.

[d] : 'diff.' is the difference between the performance and the average performance of the workers with the same tenure.

[e] : Job-level dummies and the interactions with performance.

Figure 1 Threat of Dismissal and Tenure in a Sorting Model

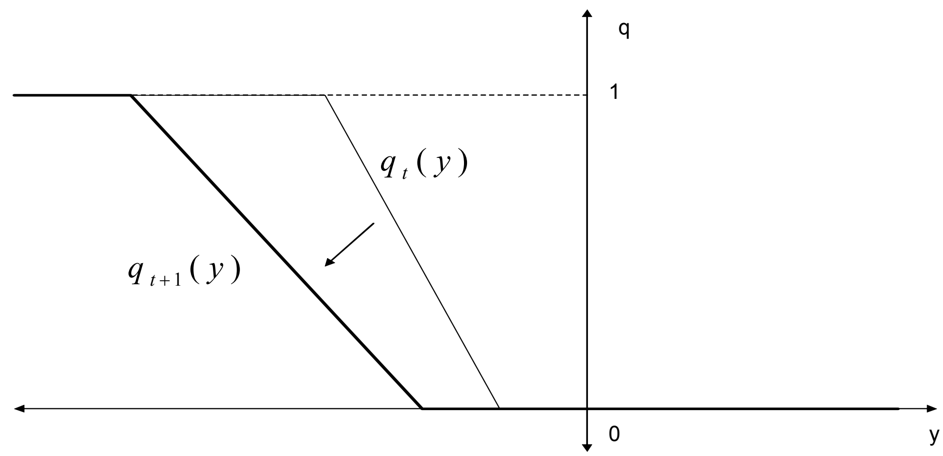


Figure 2 The Optimal Contract in an Incentive Model

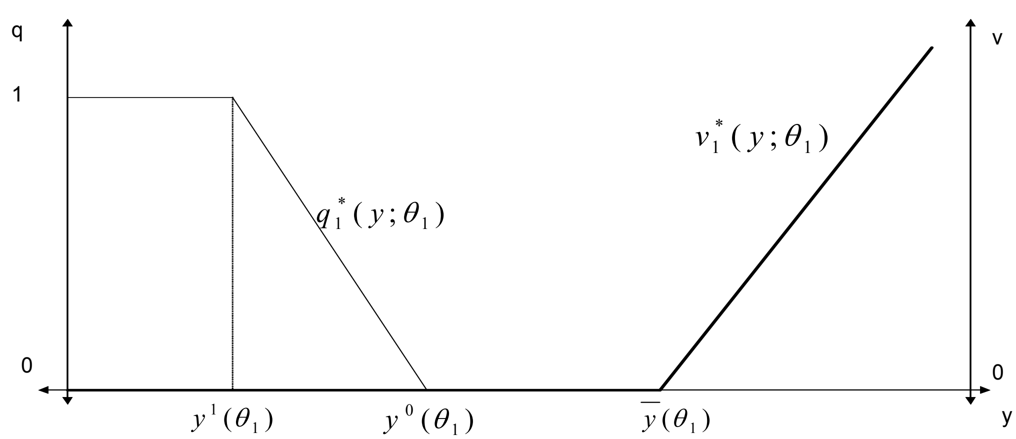
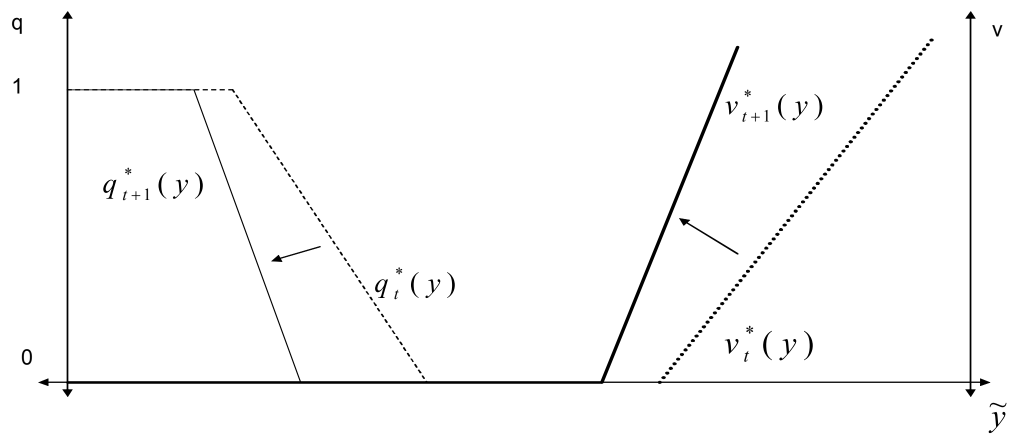
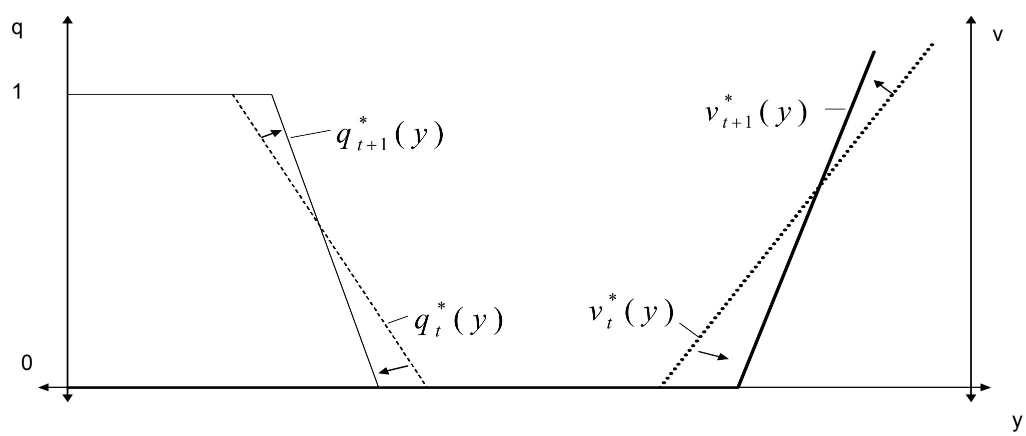


Figure 3 The Optimal Contract and Tenure in an Incentive Model



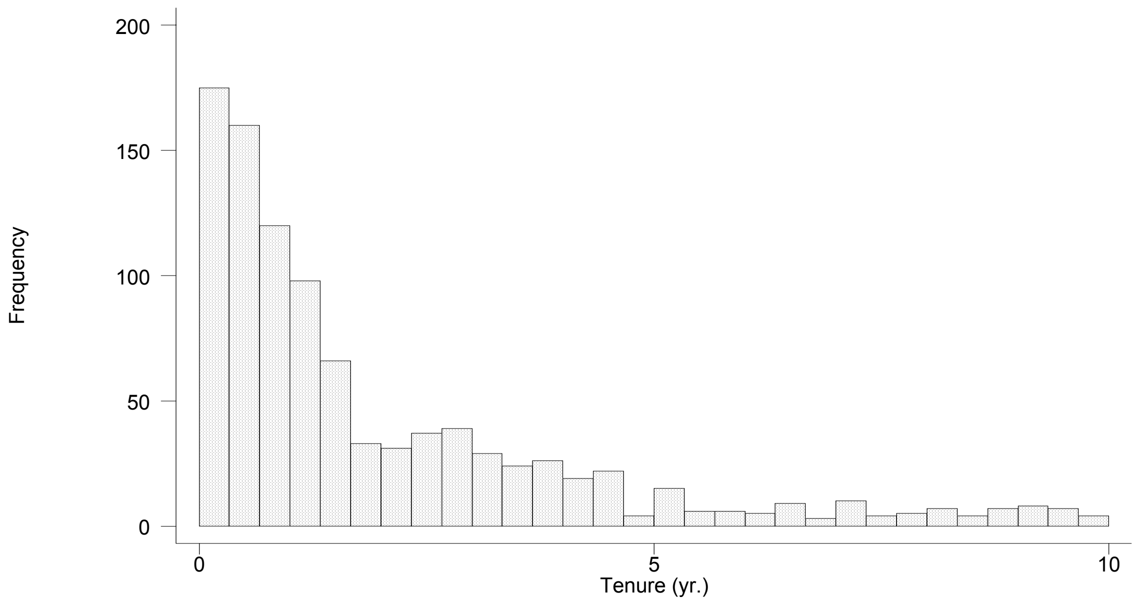
(a) In terms of mean-adjusted performance



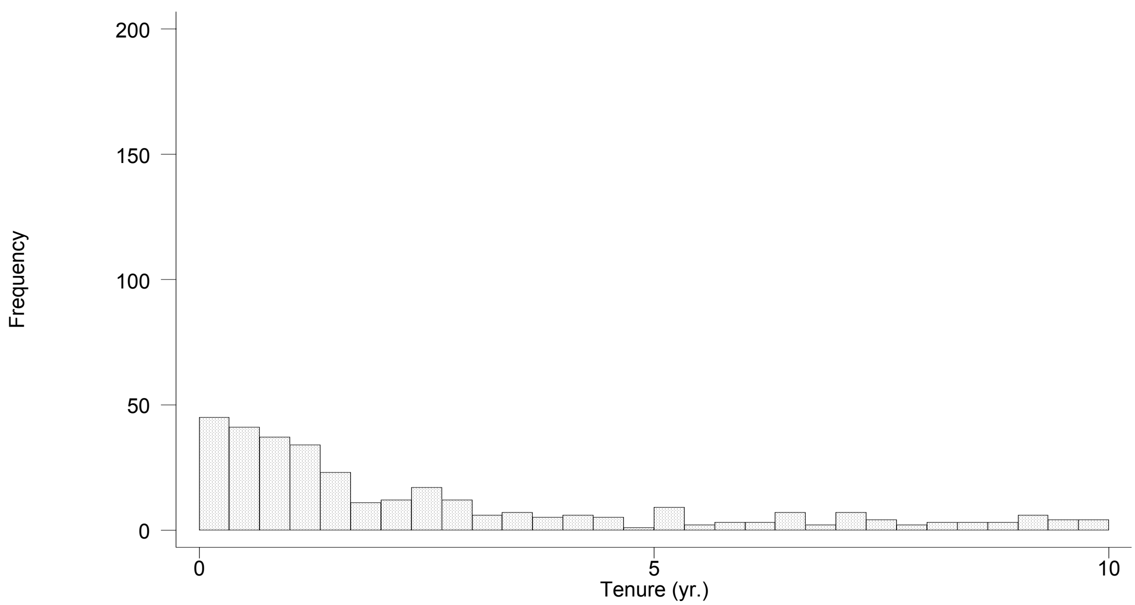
(b) In terms of (unadjusted) performance

Figure 4 Turnover and Tenure

(Histogram for tenures at the time of turnovers)



(a) All Turnovers



(b) Involuntary Turnovers

Figure 5 Tenure Distribution

(Histogram for tenures in the whole sample)

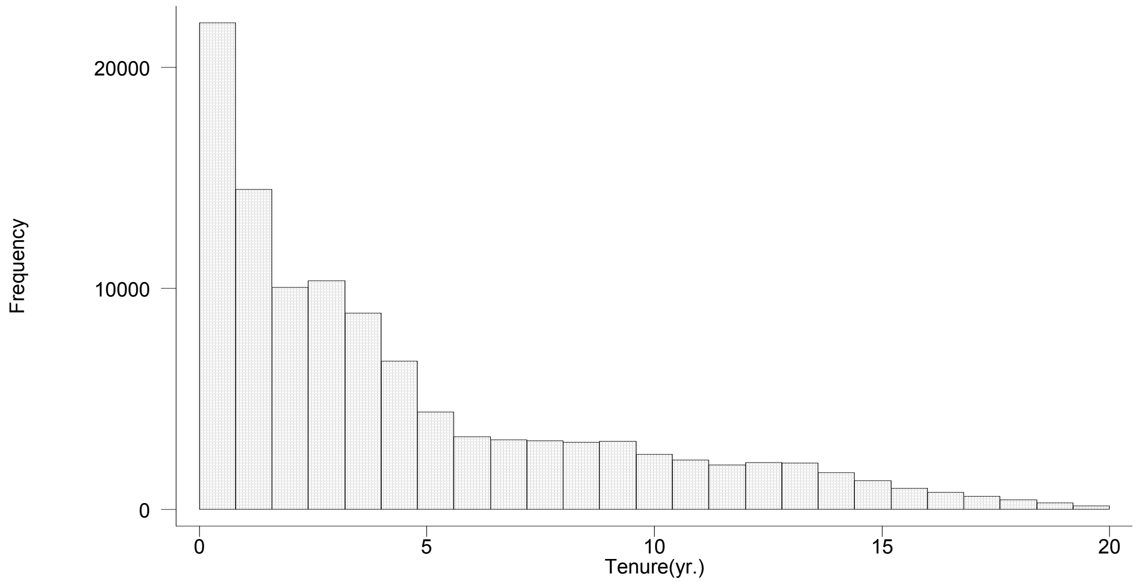
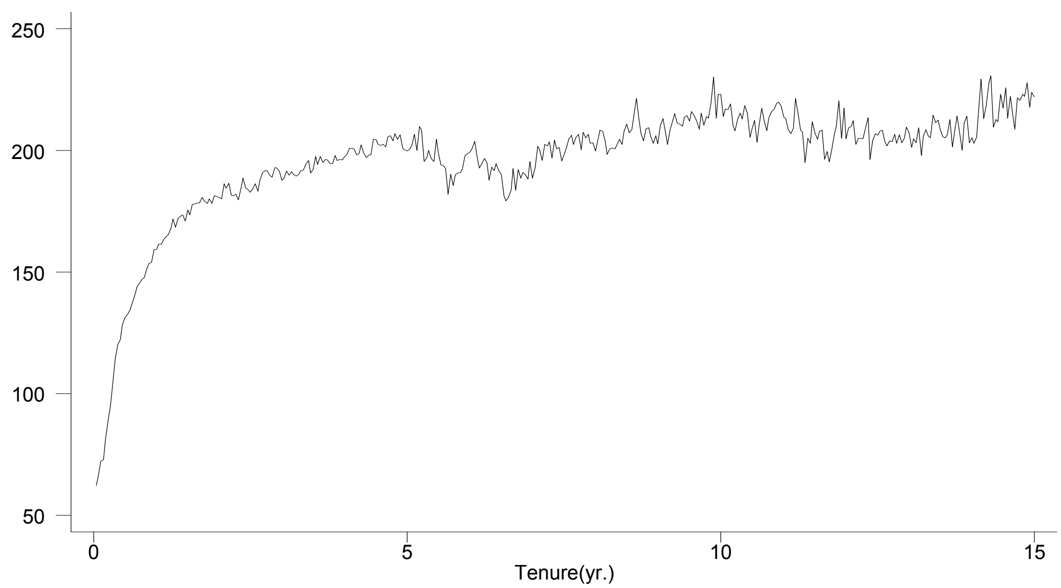
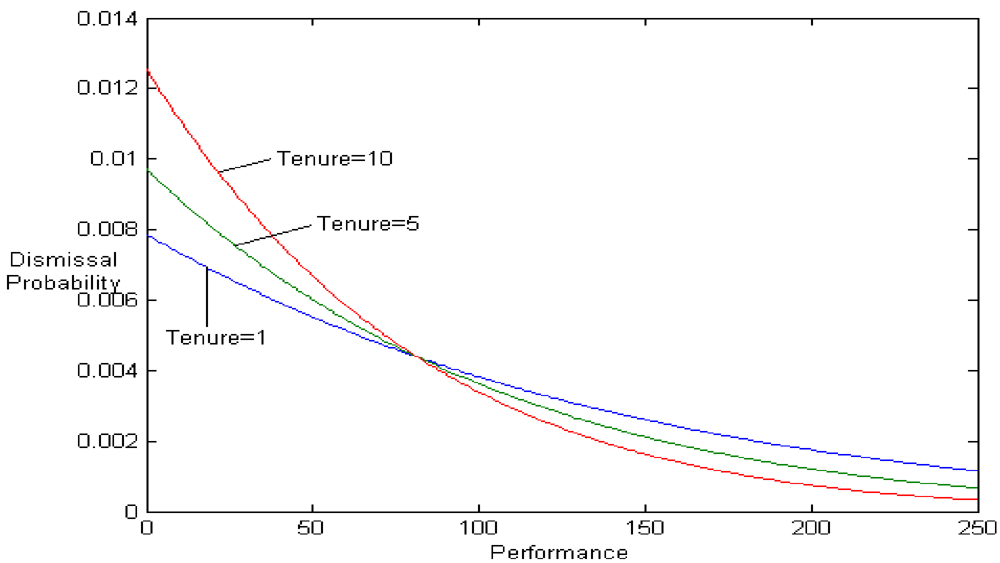


Figure 6 Productivity and Tenure

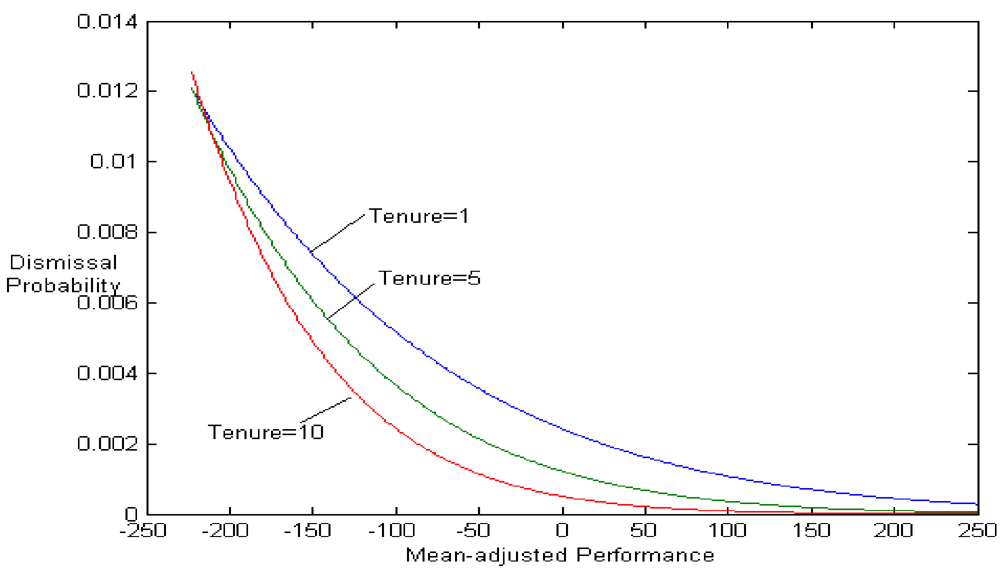


Note: The vertical axis shows the average daily performance of all workers.

Figure 7 Predicted Dismissal Probability and Tenure



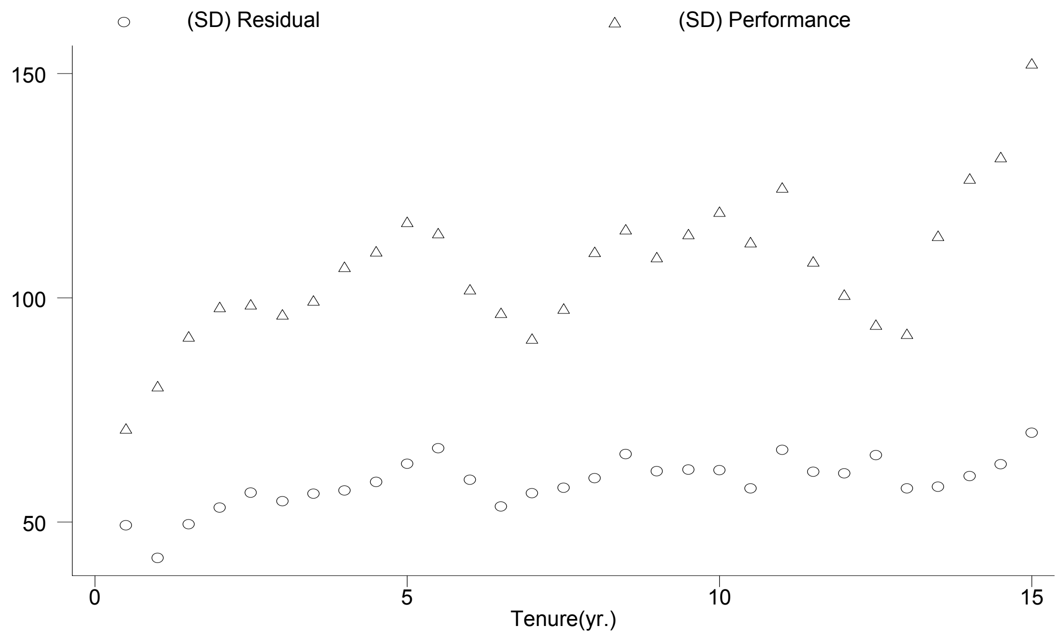
(a)



(b)

Note: The predicted values are based on the probit regression (3) in Table 5.

Figure 8 Variance of Performance and Tenure



Note: (SD) Performance is the standard deviation of the performance. (SD) Residual is the standard deviation of the residuals from the productivity regression (5) in Table 4.