Nonstationary Relational Contracts with Adverse Selection

Huanxing Yang[*†]
Department of Economics, Ohio State University

May, 2012

Abstract

We develop a model of nonstationary relational contracts in order to study internal wage dynamics. Workers are heterogeneous and each worker's ability is both private information and fixed for all time. Learning therefore occurs within employment relationships. The inferences, however, are confounded by moral hazard: the distribution of output is determined by both the worker's type and by his unobservable effort. Incentive provision is restricted by an inability to commit to long-term contracts. Relational contracts, which must be self-enforcing, must therefore be used. The wage dynamics in the optimal contract, which are pinned down by the tension between incentive provision and contractual enforcement, are intimately related to the learning effect.

JEL: C73, D82, J41, L14
Key Words: Relational Contracts; Learning; Tenure; Nonstationary; Wage Dynamics.

1 Introduction

Moral hazard pervades employment relationships. One way to alleviate the moral hazard problem is to use contingent contracts. However, the non-verifiability of workers' performance practically limits the usage of court-enforced contingent contracts. Nevertheless, if an employment relationship is repeated indefinitely, parties may rely on *relational contracts* that include both formal (court-enforced) and informal provisions. Since the informal provisions are not legally enforceable, they have to be self-enforcing – that is, each party should have no incentive to deviate from the informal provisions. This self-enforcing requirement imposes a contractual enforcement constraint on relational contracts.

There is a growing literature on relational contracts (Bull, 1987; Baker et.al, 1994; MacLeod and Malcomson, 1989, 1998; Levin, 2003). However, all of these papers focus on *stationary contracts* with contractual terms invariant with respect to the length of relationships. In reality, contractual

[*]Email: yang.1041@osu.edu. Mailing address: 448 Arps Hall, 1945 N. High St., Columbus, OH 43210.

[†]I am grateful to Jan Eeckhout, Dirk Krueger, Volker Nocke, George Mailath, Nicola Persico, Rafael Rob, Petra Todd, and especially to Steven Matthews and two anonymous referees for valuable comments. I also would like to thank the seminar participants at Iowa, OSU, LSE and SMU for their useful comments.

terms often vary with the length of relationships. The wage-tenure effect – wage increases with tenure – has been a well established stylized fact (Mincer, 1974; Becker, 1975; Jovanovic and Mincer, 1981; Topel, 1991).

The main purpose of this paper is to develop a model of *nonstationary* relational contracts to account for wage dynamics. We do so by incorporating adverse selection (heterogenous workers), which creates a learning effect: firms learn the characteristics of workers as the relationships continue. The main message of the paper is that it is the interaction between incentive provision and contractual enforcement that ties wage dynamics to the learning effect, thus making wage increasing with tenure.

More specifically, we construct a repeated principal-agent model with the following key features. First, we model a labor market as a repeated matching market, with matches constantly reshuffled. This reshuffling is partly exogenous and partly endogenous, i.e., induced by workers' or firms' decision whether to continue the current relationship. Second, workers are heterogenous. Low type workers are inherently inept, while high type workers are potentially productive but have a moral hazard problem: they choose an unobservable effort based on incentives. A worker's type is *persistent* and is his own private information. Third, a worker's output is only observable to his current employer, not to the court nor to other potential employers. Fourth, workers have limited liability. Finally, following the relational contract literature, we assume that firms cannot commit to long-term contracts; the only legally binding contracts are spot non-contingent contracts.

We focus on high-effort equilibria with high type workers exerting effort in every period. In each relationship, a relational contract specifies the conditions under which the relationship continues and wage as a function of tenure; and if either party is found to have deviated, the employment relationship is endogenously terminated. The heterogeneity among workers creates a learning effect: as a relationship continues the current employer learns the worker's type more accurately. Moreover, the learning is local or confined to the current employment relationships since workers' outputs are not observable to outsiders. To motivate high type workers to exert effort, wage must increase with tenure (at least across some tenure periods). However, the contractual enforcement constraint entails that wage cannot increase too fast with tenure; since, otherwise, senior workers would be less profitable than new workers, and firms will renege by terminating the current employment and hiring new workers. This tension between incentive provision and contractual enforcement drives the main results of the paper.

We study the conditions under which high-effort equilibria exist and the wage dynamics under the optimal contract(s) that maximize firms' expected profits. We establish that high-effort equilibria exist only if the proportion of low type workers is not too small nor too large. This implies that the presence of adverse selection might help alleviate moral hazard when firms are not able to commit to long-term contracts in a repeated matching market setting. Intuitively, the learning effect created by the presence of low type workers can alleviate the tension between incentive provision and contractual enforcement: the expected productivity of a worker increases with tenure due to the learning effect, so wage can increase with tenure without violating firms' no-reneging conditions.

If high-effort equilibria exist, then there is a unique optimal contract, under which the wage dynamics exhibits two salient features. First, wage is low and remains constant in earlier tenure periods. Second, when wage begins to increase in later tenure periods, the wage increases are intimately related to the learning effect: the wage increase between two tenure periods exactly equals the increase in the worker's expected productivity. Intuitively, since low type workers are more likely to have a short tenure, in order to minimize the informational rent to low type workers, firms try to "backload" wages: pay low wages in earlier tenure periods and use wage increases in later tenure periods to provide incentives for high type workers. However, the contractual enforcement constraint limits firms' ability to backload wages: the wage increases cannot exceed the learning effect. As a result, in the optimal contract the wage increases in later tenure periods are tied to the learning effect. One interesting point is that, although learning is completely confined to current matches in our model, the wage increases are tied to the learning effect. This implies that, even without market competition, wages being tied to workers' expected productivities can be generated by *internal wage dynamics*.

The rest of the paper is organized as follows. The next subsection discusses the related literature. Section 2 sets up the model. Some preliminary analysis is offered in Section 3 and Section 4. Section 5 studies the existence of high-effort equilibria and optimal contracts. Section 6 discusses separating contracts and Section 7 concludes. All of the proofs can be found either in the Appendix or in an online appendix.

Related literature This paper is related to several strands of literature. The first strand of literature is relational contracts (Bull, 1987; Baker et.al, 1994; MacLeod and Malcomson, 1989, 1998; Levin, 2003). As mentioned earlier, this paper differs from those papers in that we study nonstationary contracts. Actually, except for Levin (2003) none of those papers incorporate adverse selection. In Levin (2003) the worker has hidden information. However, the worker's type is not persistent. Recently, Fong and Li (2010) incorporate limited liability into a model of relational contracts. In their model, there is only moral hazard, and the element of hidden information is not present. MacLeod and Malcomson (1988) study a dynamic employment model with both moral hazard and adverse selection of persistent types. They show that the equilibrium contract consists of a hierarchy of ranked jobs, with workers producing satisfactory performance in low ranked jobs having the potential to be promoted to high ranked jobs in the future. In their model, a worker's output is deterministic given his effort. Thus, firms' learning about workers' type in their model is quite different from that in the current model, which leads to different contract dynamics.

Several recent papers study nonstationary relational contracts, focusing on issues different from the current paper's. Halac (2011) studies the dynamics of a relationship in which the principal has persistent private information regarding her own outside option. Chassang (2010) considers a dynamic cooperation game in which one agent has private information about which actions are productive and the other agent learns the set of productive actions over time. His main focus is on how parties figure out and settle on the details about cooperation. Thomas and Worrall (2010) study a long-term relationship consisting of two agents who undertake investments in each period. The investments are relationship-specific, which creates a hold-up problem. They address the issues of how investment should be structured and how surplus should be divided over time in order to alleviate the hold-up problem.

Except for MacLeod and Malcomson (1989, 1998), other previously mentioned papers on relational contracts restrict attention to one-principal-one-agent settings, thus both parties' outside options are exogenously given. In our model, workers and firms live in a repeated matching market, thus both workers' and firms' outside options in current relationships are endogenously determined. A recent paper by Board and Meyer-ter-Vehn (2011) also considers relational contracts in a market setting. Specifically, in their model, firms offer different incentive contracts and employed workers can search for better jobs. They study how changes in on-the-job search affect equilibrium distribution of contracts.

The second strand is the literature on the wage-tenure effect. There are two existing non-contractual approaches to explain the wage-tenure effect. Neoclassical human capital theory (Becker, 1962; Hashimoto, 1981) argues that wage increases with tenure because individual workers' productivities increase with firm-specific human capital accumulation. The second one is Jovanovic's (1979) matching model with learning. In his model, within each individual match, a firm and a worker symmetrically learn the quality of the match. Moreover, low quality matches endogenously break up and only high quality matches remain. This learning effect, combined with endogenous separation, leads to the wage-tenure effect.

Our paper differs from Jovanovic (1979) in two aspects. First, our paper models the dynamic contracting problem explicitly. Second, in Jovanovic's model, wages being tied to the learning effect is due to the market's competition for workers, as workers' past performance is commonly observed. In our model, learning is confined within the current matches, and the wage-tenure effect results from *internal* wage dynamics. Felli and Harris (1996) endogenize the wage determination in Jovanovic's model, but they confine it to a setting in which two firms are competing for the service of a worker over time. In their model, the wage-tenure effect exists only if there is a learning externality: learning in the current match also provides information about the workers' productivity in the alternative match.[1]

In a pure moral hazard model, Lazear (1979) considers the increasing wage profile as a contrac-

[1] Burdett and Coles (2003) study the wage-tenure effect in a job search framwork. Their main foucs is to separate the wage-tenure effect from wage growth due to searching for better jobs. And they also assume that firms can commit to long-term contracts.

tual device to prevent workers from shirking. However, he assumes that firms are able to commit to long-term contracts. Moreover, his model cannot pin down the wage dynamics, as there are many increasing wage-tenure profiles that can prevent workers from shirking. Harris and Holmstrom (1982) develop a model of wage dynamics based on symmetric learning and insurance concerns. Their model is more relevant in accounting for the relationship between wages and general working experience. Moreover, they also assume that firms can commit to long-term contracts.

The third strand of related literature studies how cooperation can be achieved in repeated matching markets (Dutta, 1993, Ghosh and Ray, 1996; Kranton, 1996; Rob and Yang, 2010). In particular, Ghosh and Ray (1996) and Rob and Yang (2010) show that the presence of "bad" type agents can discipline opportunists to adopt cooperative behavior.[2] In those models, however, there are no contracts, hence providing no implications about contract dynamics. Moreover, in both models, players perfectly learn their partners' type after the first period of interaction. In our model, monitoring is imperfect, so learning is gradual unless separating contracts are offered.[3]

2 The Model

There is a continuum of firms with measure 1, and each firm has exactly one job vacancy.[4] Correspondingly, there is a continuum of workers with measure 1. All workers and firms are risk-neutral, live forever and share the same discount factor δ. Time is discrete, indexed by $PT = 1, 2, ...$ In each period, workers and firms are matched to engage in production. Each existing match will continue in the next period with probability $\rho \in (0, 1)$ and break up with probability $1 - \rho$ for exogenous reasons. A match can also be dissolved endogenously if either party in the current match decides to leave the match. All the agents in dissolved matches enter into the unmatched pool, and they are randomly paired at the beginning of the next period. The time line will be specified shortly. Note that workers and firms are of equal measure, so each agent is guaranteed a match at the beginning of each period.[5]

The stage output y for a match is either 0 or 1, and the value of output y is y. Workers are of two types: high type H and low type L. The measure of L type workers is $\beta \in [0, 1]$, and the measure of H type is $1 - \beta$. A worker's type is fixed for all time and is his own private information. The two types of workers differ in productivity: H type workers have an option to choose a high

[2] Mailath and Samuelson (2001) establish that reputational concerns can also be generated by a high type firm's incentive to differentiate itself from low types. But their model focuses on reputation and only studies the one-firm case.

[3] This paper is also loosely related to the following papers. For adverse selection in labor markets , see Greenwald (1986). For information asymmetry between the current employer and alternative firms, see Waldman (1984b) and Bernhardt (1995). For symmetric and public learning in labor market contexts, see Holmstrom (1999) and Farber and Gibbons (1996).

[4] The main results of the paper still hold as long as each firm has a finite number of job vacancies.

[5] If the measures of workers and firms are not equal, then the long side of the market will have matching friction. For this direction of research, see MacLeod and Malcomson (1989, 1998).

effort $\bar{e} > 0$ or a low effort 0; L type workers are inept and can only exert low effort 0. One interpretation is that, even if an L type worker exerts the high effort $\bar{e}$, the distribution of output is the same as if he were to exert effort 0. The cost of effort $\bar{e}$ is $c > 0$ and the cost of effort 0 is 0. A worker's effort is not observable.

Output y only depends on the effort level. Specifically,

$$Pr\{y = 1|e\} = \begin{cases} 1 & \text{if } e = \bar{e} \\ p \in (0,1) & \text{if } e = 0 \end{cases} .$$

This assumption implies that monitoring is imperfect, in the sense that output does not perfectly reveal a worker's effort. It also implies that a H type worker who exerts 0 effort is the same as an L type worker in terms of productivity. We assume $1 - p > c$, so the efficient action for H workers is $\bar{e}$. A worker's output y is observable to the worker and his current employer, but not to the court or to other market participants. Thus, court-enforced contracts that are contingent on y are not feasible, and there is no information flow between matches. We assume that a worker's previous employment history is not observable to firms,[6] and a firm's previous employment history is not observable to workers, either. If a worker is not employed in one period, he gets a reservation utility 0 in that period regardless of his type. Since $p > 0$, it is efficient for both types of workers to be employed in each period. Similarly, if a firm does not employ a worker, its profit in that period is 0.

Firms are not able to commit to long-term contracts. The only legally binding contracts are spot contracts, which specify a fixed wage payment w_t. Here t denotes the tenure period (starting from 1), which is the number of periods that a worker has been matched with the current firm. A firm may also offer its worker a discretionary bonus b_t in tenure period t which the firm promises to pay if and only if $y_t = 1$. At the beginning of employment, a firm and its matched worker agree upon how the payments are going to evolve as the relationship continues. We name the agreed-upon payment plans $\{w_t, b_t\}$ as contracts. There are two kinds of payment plans: *pooling contracts* and *separating contracts*. Under pooling contracts both types of workers have the same payment plan, while under separating contracts firms offer two contracts and let workers self-select in tenure period 1. We will focus mainly on pooling contracts, which are denoted as $\{w_t, b_t\}$. We will refer to pooling contracts simply as "contracts" unless further clarification is necessary. Separating contracts will be considered in Section 6. Finally, workers are subject to limited liability, that is, $w_t \geq 0$ for all t.

Figure 1 specifies the time line within a period. At the beginning of a period, unmatched workers and unmatched firms are paired randomly. In each newly formed match, a spot contract for tenure period 1, (w_1, b_1), is offered, and the worker decides whether to accept or to reject the

[6] This is a simplifying assumption, which makes workers in the unmatched pool homogenous in appearence.

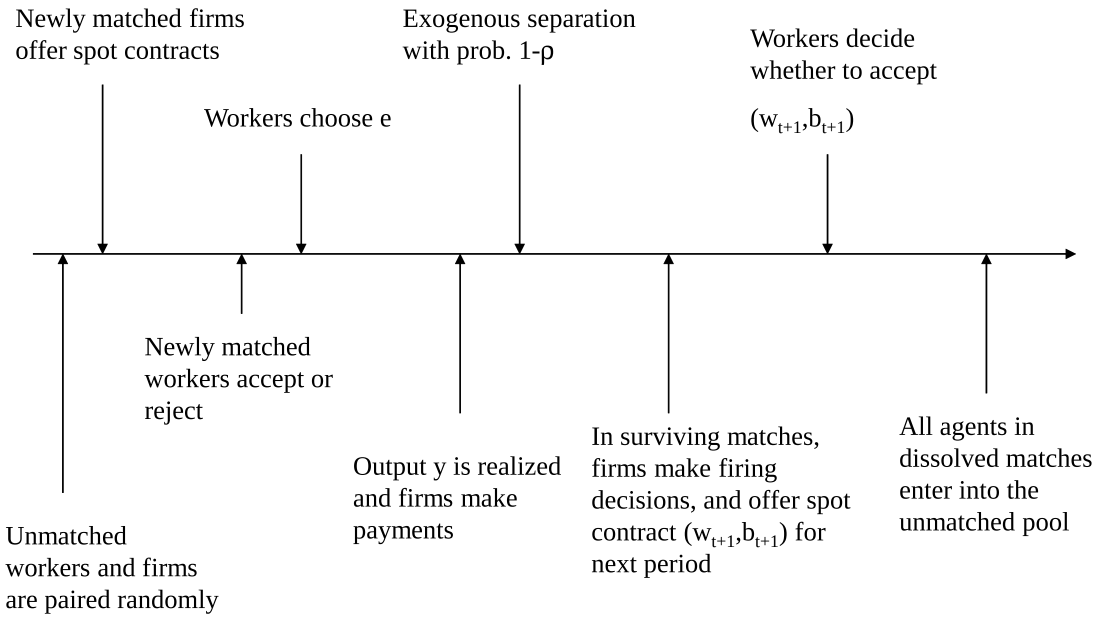

Figure 1: Time Line of a Typical Period

offer. If a worker rejects the offer, he leaves the match and collects reservation utility 0 in that period. Then, among all employed workers, H type workers choose their effort level. Afterwards, output y (in each match) is realized and workers are paid. Then, exogenous separation occurs in existing matches with probability $1-\rho$. In each surviving match, firms make firing decisions. If a firm wants to retain the worker, it offers a spot contract (w_{t+1}, b_{t+1}) for the next period, and the worker decides whether to accept the offer. A match is dissolved endogenously if the firm fires the worker or the worker rejects the firm's offer. All agents in dissolved matches enter into the unmatched pool. Finally, the next period begins.

3 Preliminary Analysis

We adopt symmetric perfect public equilibrium (SPPE) as our solution concept. By symmetry we mean that all firms adopt the same strategy and each type of worker also adopts the same strategy. Public strategies require that each agent's strategy only depends on the public history within the current relationship, since the previous employment history is not observable.[7] Let W_t be the total wage payment actually made in tenure period t. The public history of a relationship that has lasted for t tenure periods can be denoted as $h^t = (w_1, b_1, y_1, W_1; ...; w_t, b_t, y_t, W_t)$. A (behavior) strategy for an H type worker, σ^H, is decision rules about whether to accept the spot contract and what effort level to choose, both as a function of public history. A strategy for an L type worker, σ^L,

[7] This implies that each relationship is played out in the same way in equilibrium.

is a decision rule about whether to accept the spot contract as a function of public history. A strategy for a firm, σ^f, specifies whether to fire the worker and what spot contract to offer, both as a function of public history. A relational contract, which is a complete plan for a relationship, consists of a strategy profile $\sigma = (\sigma^H, \sigma^L, \sigma^f)$. Denote $\phi(h^{t-1})$ as a firm's belief that its worker is of H type, given history h_{t-1}.

There is always a trivial equilibria in which H type workers always exert 0 effort, firms always offer 0 wage, workers always accept nonnegative wage offers, and firms always fire their current workers. Given that H type workers always exert 0 effort, firms have no incentive to offer positive wages and firing decisions become irrelevant. In this equilibrium of zero-wage contracts, each firm gets a per-period profit p. Such non-reputational equilibrium is not the focus of this paper. Recall that the efficient outcome is for all workers to be employed and H type workers to exert high effort in each period. We call equilibria with this outcome *high-effort equilibria*. These equilibria are the primary focus of this paper.

A necessary condition for a high-effort equilibrium is that H type workers should have an incentive to exert high effort $\bar{e}$ in each period (*no-shirking conditions*). To effectively prevent shirking, we restrict attention to the following trigger strategy: a firm retains its worker if and only if the worker produces $y_t = 1$ in each previous tenure period and fires the worker immediately if $y_t = 0$.

Given that only fixed-wage spot contracts can be legally enforced, another necessary condition for a high-effort equilibrium is that firms have no incentive to renege (*no-reneging conditions*). Specifically, there are three kinds of reneging. First, firms can renege on bonus b_t by not paying b_t when $y_t = 1$. Second, a firm's spot contract offer in tenure period t can be different from the equilibrium payment plan $\{w_t, b_t\}$, which was implicitly agreed upon by all parties. Third, a firm can fire a worker even if the worker always produces $y_t = 1$ in the relationship (recall that the firms' trigger strategy specifies that a worker is retained if he always produces $y_t = 1$ in the relationship). Intuitively, if senior workers are less profitable than new workers, then firms may fire senior workers regardless of their performance. To effectively deter firms' reneging, we restrict attention to the following trigger strategy: a worker stays in his current firm if and only if the firm always pay bonus b_t and the spot contracts have always followed the equilibrium plan $\{w_t, b_t\}$; otherwise, he quits immediately.

We focus on trigger strategies because they provide the severest punishment for the deviating party, thus making high-effort equilibria easier to sustain. A trigger strategy is clearly a best response for firms since only L type workers produce $y = 0$ on the equilibrium path. A trigger strategy is also a best response for workers (both types), if we assume that workers hold the most pessimistic belief off the equilibrium path. More specifically, once a firm deviates from the equilibrium payment plan $\{w_t, b_t\}$, the worker holds the belief that the firm will offer the lowest possible wage (0) in all future spot contracts if the relationship continues. Given this belief, it is a best response for the worker, regardless of his type, to quit immediately after the firm deviates.

The next simplifying step is that, without loss of generality, we can restrict attention to fixed wage contracts $\{w_t\}$. That is, it is without loss to set $b_t = 0$ for all t. The underlying reason is that, for any contract that includes bonus payments, $\{w_t, b_t\}$, there is always a corresponding payoff-equivalent fixed wage contract $\{w'_t\}$.[8] Therefore, we will only consider fixed wage contracts $\{w_t\}$ hereafter.

With fixed wage contracts, we do not need to worry about firms' reneging on bonus payments. Note that the workers' trigger strategy effectively deters firms' reneging of the second category: a firm's spot contract offers will always follow the equilibrium $\{w_t\}$ if it wants to retain the worker. Therefore, we only need to worry about firms' reneging of the third category. As a result, firms' no-reneging conditions boil down to the condition that firms always have an incentive to retain a worker who has always produced $y_t = 1$ in previous periods of the relationship.

Under trigger strategies, tenure period t is a sufficient statistic of the previous public history. A worker in tenure period t means that $y_j = 1$ for all $j \leq t-1$ in the current firm and the wage offers have followed $\{w_t\}$ so far. At any physical time PT, H type workers will be in different tenure periods because of exogenous separation. Type L workers are also in different tenure periods because of imperfect monitoring. Define x_t (β_t) as the population of H (L) type workers who are in tenure period t. We restrict attention to a *stationary state*, that is, the distributions of the types of workers in different tenure periods $\{x_t\}$ and $\{\beta_t\}$ are invariant with respect to physical time PT.[9] In the stationary state, $\beta_t = (\rho p)^{t-1}\beta_1$. Summing up β_t and using the fact that the total population of L type workers is β, we get $\beta_1 = (1-\rho p)\beta$. Similarly, one can get $x_t = \rho^{t-1}x_1$ and $x_1 = (1-\rho)(1-\beta)$ in the stationary state.[10] The following definition formally summarizes the equilibrium conditions for high-effort equilibria.

Definition 1 *A (trigger strategy) high-effort equilibrium with (pooling) contract $\{w_t\}$ satisfies: (i) all workers accept offers in tenure period 1, and all firms have incentives to employ new workers (participation constraints), (ii) H type workers will exert high effort $\overline{e}$ in each period (no-shirking conditions), (iii) firms always retain a worker who always produces $y_t = 1$ in the relationship (no-reneging conditions), (iv) no firm has an incentive to unilaterally deviate to offering the zero-wage contract.*

Note that, in the above definition of equilibrium, there is no need to worry about firms deviating to other contracts except for the zero-wage contract. This is because an equilibrium contract $\{w_t\}$

[8] A formal proof of this claim can be found in an earlier version of the paper. The idea is that any bonus b_t can be incorporated into the fixed wage payment of the next tenure period, w_{t+1}, without affecting the expected payoff for each party. Levin (2003) establishes that focusing on stationary bonus contracts is without loss of generality. The difference is that, in his model, there is no persistent type, which essentially yields a stationary environment in terms of contracting.

[9] We assume that the economy settles into the stationary state in the first physical time period.

[10] On the equilibrium path, H type workers turn over only because of exogenous separation.

is like a social norm: if a firm offers any contract other than $\{w_t\}$, workers will exert zero effort and quit immediately in the next period (trigger strategy supported by workers' worst belief after observing any off-equilibrium contracts). As a result, the most profitable deviation for firms in terms of offering other contracts is to offer the zero-wage contract. The equilibrium concept adopted in MacLeod and Malcomson (1989) has the same flavor.

One important observation is that a firm learns its worker's type as tenure period t increases. Under trigger strategies, a firm's initial belief in tenure period t, $\phi(h^{t-1})$, can be simply denoted as ϕ_t. Recall the assumption that workers' previous employment history is not observable. This leads to a common initial belief ϕ_1 about all workers in the unmatched pool. Specifically,

$$\phi_1 = \frac{x_1}{x_1 + \beta_1} = \frac{(1-\rho)(1-\beta)}{(1-\rho)(1-\beta) + (1-\rho p)\beta}. \tag{1}$$

From (1), it is evident that ϕ_1 is decreasing in β. Firms update their beliefs according to Bayes' rule as follows:

$$\phi_t = \frac{\phi_1}{\phi_1 + p^{t-1}(1-\phi_1)}. \tag{2}$$

Observing (2), we see that ϕ_t only depends on tenure period t and is increasing in t. In other words, a firm's belief about its worker becomes more optimistic as the relationship continues since low type workers are gradually weeded out. To abuse notation somewhat, we denote y_t as a worker's expected output in tenure period t. That is, $y_t \equiv \phi_t + p(1-\phi_t)$. Note that y_t is increasing in t since ϕ_t is. By (2), we can see that $y_t = \phi_t/\phi_{t+1}$.

Define U_t (U_t^L) as the equilibrium discounted payoff of a type H (L) worker who is in tenure period t. The recursive value functions are:

$$\begin{aligned} U_t &= (w_t - c) + \delta[\rho U_{t+1} + (1-\rho)U_1]; & (3)\\ U_t^L &= w_t + \delta[p\rho U_{t+1}^L + (1-p\rho)U_1^L]. & (4) \end{aligned}$$

Define U_t^d as the discounted payoff of a type H worker who is in tenure period t and shirks only in that period,

$$U_t^d = w_t + \delta[p\rho U_{t+1} + (1-p\rho)U_1]. \tag{5}$$

Similarly, define V_t as a firm's equilibrium discounted profit who currently matches with a tenure-period t worker:

$$V_t = (y_t - w_t) + \delta[\rho y_t V_{t+1} + (1-\rho y_t)V_1]. \tag{6}$$

Let V_t^d be a firm's discounted profit which currently matches with a tenure-period t worker and reneges in that period. As discussed earlier, the only reneging we need to consider is that the firm fires its worker who has produced $y_j = 1$ for any $j \leq t$. Thus,

$$V_t^d = (y_t - w_t) + \delta V_1. \tag{7}$$

Note that all the value functions are nonstationary due to the gradual learning effect.

Now the no-shirking conditions can be explicitly written as:

$$\begin{aligned} & U_t - U_t^d \geq 0 \;\Leftrightarrow \delta\rho(1-p)[U_{t+1} - U_1] \geq c \text{ for any } t \geq 1 \\ \Leftrightarrow \quad & U_t - U_1 \geq \widehat{c} \;\text{ for all } t \geq 2, \text{ where } \widehat{c} \equiv \frac{c}{\delta\rho(1-p)}. \end{aligned} \tag{8}$$

Inequality (8) says that, to prevent H type workers from shirking, the equilibrium discounted payoff in later tenure periods relative to that in the first tenure period has to be big enough. This implies that, in general, w_t has to be nondecreasing in t. Similarly, firms' no-reneging conditions can be written as:

$$\begin{aligned} & V_t - V_t^d \geq 0 \Leftrightarrow V_{t+1} - V_1 \geq 0 \text{ for all } t \\ \Leftrightarrow \quad & V_t - V_1 \geq 0 \text{ for all } t \geq 2. \end{aligned} \tag{9}$$

According to (9), to prevent a firm from reneging, its equilibrium discounted payoff when matched with a senior worker should be no smaller than that when matched with a new worker. That is, senior workers cannot be less profitable than new workers. Note that both no-shirking conditions and no-reneging conditions consist of an infinite number of constraints. To ease exposition, we refer to a contract $\{w_t\}$ that satisfies the no-shirking conditions (8) and no reneging conditions (9) as a *self-enforcing contract.*

Workers' participation constraints require $U_t \geq 0$ and $U_t^L \geq 0$ for all t. However, given that H type workers can always mimic L type workers, the no-shirking conditions (8) imply $U_t \geq U_t^L$ for all t. Thus, workers' participation constraints boil down to $U_t^L \geq 0$ for all t. But given that $w_t \geq 0$ due to limited liability, by (4) $U_t^L \geq 0$ for all t is always satisfied. Firms' participation constraints require $V_t \geq 0$ for all t. But given the no-reneging conditions (9), $V_1 \geq 0$ is sufficient. Note that, if a firm always offers the zero-wage contract, it can earn a discounted payoff $p/(1-\delta)$. Thus requirement (iv) of high-effort equilibria is equivalent to $V_1 \geq p/(1-\delta) > 0$, which ensures that firms' participation constraints are satisfied as well. The following lemma summarizes the above analysis.

Lemma 1 *A high-effort equilibrium exists if and only if there is a contract $\{w_t\}$ such that: (i) $\{w_t\}$ is self-enforcing, or satisfies (8) and (9), and (ii) V_1 under $\{w_t\}$ is bigger than $p/(1-\delta)$.*

As multiplicity of equilibria is typical in repeated games, in our model there might be multiple high-effort equilibria associated with different contracts. Note that all high effort equilibria yield the same social surplus since the efficient outcome is implemented in each period. Among all possible high-effort equilibria, we are interested in the equilibrium in which firms' discounted payoff, V_1, is maximized. In particular, we refer to a contract $\{w_t\}$ in the high effort equilibrium that maximizes V_1 as an *optimal contract.*

The rest of the paper will focus on the following issues. The first one is identifying the conditions under which high-effort equilibria exist. The second one is characterizing optimal contracts.

4 A Certain Class of Contracts

We have two major difficulties in our analysis. First, equilibrium conditions (8) and (9) are involved with two sets of an infinite number of constraints. Second, there is too much freedom in the design of contracts, which consist of an infinite sequence of wages. Fortunately, we are able to show, later in this paper, that optimal contracts must belong to a certain class of contracts that we call quasi-monotonic contracts. Let π_t be a firm's expected profit in tenure period t: $\pi_t \equiv y_t - w_t$.

Definition 2 *A contract $\{w_t\}$ is **nondecreasing** if w_t is nondecreasing in t. Let T be the first tenure period such that wage is strictly positive in a nondecreasing contract $\{w_t\}$. A nondecreasing contract $\{w_t\}$ is **quasi-monotonic** if either (i) for any $t \geq T$, $\pi_{t+1} \geq \pi_t$, or (ii) for any $t > T$, $\pi_{t+1} \geq \pi_t$, and $\pi_T > \pi_{T+1}$ and $\pi_T \geq (1-\delta)V_1$.*

Quasi-monotonic contracts have two properties. First, wages are nondecreasing in tenure. Second, roughly speaking, firms' stage profits are nondecreasing in tenure. The only exception is that firms' stage profits might be decreasing from tenure period $T-1$ to tenure period $T+1$.

Proposition 1 *If there is a self-enforcing contract $\{w_t\}$, then there is a self-enforcing quasi-monotonic contract $\{w'_t\}$. Moreover, firms' expected (discounted) profits are the same under the two contracts, $V_1 = V'_1$, and the no-shirking conditions do not bind for any $t \geq 2$ under $\{w'_t\}$.*

The proof of Proposition 1 is by construction, and it can be found in an online appendix. Proposition 1 implies that, without loss of generality, we can focus on quasi-monotonic contracts when we study the existence of high-effort equilibria and optimal contracts. To understand Proposition 1, note that the existence of high-effort equilibria hinges on the tension between high type workers' no-shirking conditions (incentive provision) and firms' no-reneging conditions (contract enforcement): while the former requires that wage increases fast enough to provide incentive to high type workers, the latter puts an upper bound on the speed at which wage increases to prevent firms from reneging. For quasi-monotonic contracts, wages being nondecreasing in tenure makes the no-shirking conditions easy to satisfy, and firms' stage profits being nondecreasing in tenure (except from tenure period $T-1$ to $T+1$) makes firms' no-reneging conditions easy to satisfy.

The following lemma specifies self-enforcing quasi-monotonic contracts.

Lemma 2 *Under a quasi-monotonic contract $\{w_t\}$, the no-shirking conditions (8) become*

$$U_2 - U_1 > \widehat{c} \Leftrightarrow \sum_{j=1}^{\infty} (\delta\rho)^{j-1}(w_{j+1} - w_j) \geq \widehat{c}, \tag{10}$$

and the no-reneging conditions (9) become: $V_T \geq V_1$ ($V_{T+1} \geq V_1$) if $\pi_{t+1} \geq \pi_t$ for any $t \geq T$ (if $\pi_{t+1} \geq \pi_t$ for any $t > T$ and $\pi_T > \pi_{T+1}$).

Actually, we can go one step further by showing that optimal contracts must be quasi-monotonic.

Proposition 2 *(i) If a quasi-monotonic contract is self-enforcing, but the no-shirking condition (10) is not binding, then it cannot be optimal; (ii) If a self-enforcing contract $\{w_t\}$ is not quasi-monotonic, then it cannot be optimal.*

To understand part (ii) of Proposition 2, note that, in optimal contracts, firms try to minimize the informational rent to low type workers. To motivate high type workers wages have to be strictly positive at some point. Since low type workers can mimic high type workers with some success, they will enjoy informational rent. Given that low type workers are more likely in earlier tenure periods, to reduce the informational rent firms have an incentive to keep wages as low as possible in earlier tenure periods and use wage increases in later tenure periods to motivate high type workers. That is, firms have incentives to backload wages as much as possible. However, firms' ability to backload wages is restricted by the no-reneging conditions as backloading wages will necessarily make hiring new workers more profitable than retaining workers in later tenure periods. Intuitively, if a contract is not quasi-monotonic, then it does not backload wages enough. Specifically, if wages are not nondecreasing (strictly decrease between some tenure periods), then the wages in some earlier tenure periods are too high and firms can backload wages further. Similarly, if firms stage profits are not nondecreasing (strictly decrease between some tenure periods), it implies that wages in some later tenure periods are not high enough,[11] and firms have room to backload wages further. For these reasons, optimal contracts must be quasi-monotonic.

From now on, we will focus on quasi-monotonic contracts. Formally, we define the following programming problem [PP]:

$$\begin{aligned} \max_{\text{quasi-montonic } \{w_t\}} & V_1 \\ \text{subject to} \quad : \quad & \text{(i) the no-shirking condition (10) holds,} \\ \text{(ii) } V_T \geq \quad & V_1 \text{ and } V_{T+1} \geq V_1, \\ \text{(iii) } V_1 \geq \quad & p/(1-\delta). \end{aligned} \tag{PP}$$

By Lemmas 1 and 2 and Proposition 1, a high-effort equilibrium exists if and only if there is a quasi-monotonic contract $\{w_t\}$ satisfying the three constraints of [PP], and optimal contracts are solutions to [PP]. Note that a bigger V_1 makes constraint (iii) more easily satisfied. Therefore, the existence of high-effort equilibria boils down to the condition that [PP] has a solution. Moreover, by part (ii) of Proposition 2, optimal contracts must be solutions to [PP].

[11] Suppose $\pi_{j+1} < \pi_j$. To satisfy firms' no-reneging condition in tenure period $j+1$ ($V_{j+1} \geq V_1$), this implies that the weighted averge of firms' stage profits in tenure periods after $j+1$ have to be relatively bigger (relative to the case that $\pi_{j+1} \geq \pi_j$). This implies that wages in tenure periods after $j+1$ are relatively low.

5 Existence of High-effort Equilibria and Optimal Contracts

Inspecting the programming problem [PP], two observations are in order. First, $w_1 = 0$ in optimal contracts since what matters for the no-shirking and the no-reneging conditions is the wage increases across tenure periods, Δw_t. Second, by part (i) of Proposition 2, in optimal contracts the no-shirking condition (10) must be binding.

Lemma 3 *If the programming problem [PP] has a solution, it also has a solution of the following form: (i) $\pi_t = \pi_{T+1}$ for any $t > T+1$, (ii) $V_{T+1} = V_1$ and*

$$\pi_{T+1} = (1-\delta)V_1 = \frac{\sum_{t=1}^{T}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}\pi_t}{\sum_{t=1}^{T}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}}. \tag{11}$$

Moreover, optimal contracts must have the above form.

Lemma 3 provides a sharper characterization of the optimal contracts. Specifically, after tenure period $T+1$ (the second tenure period in which the wage is strictly positive), firms' stage profits are constant, which implies that the increase in wage exactly matches the increase in workers' expected productivity. Moreover, firms' no-reneging conditions are binding after tenure period $T+1$, which implies that firms' (constant) stage profit after tenure period $T+1$ is a weighted average of the stage profits in the first T tenure periods (equation (11)). Intuitively, Lemma 3 results from firms' incentive to backload wages. To minimize informational rents to low type workers, it is always better for firms to minimize wages in earlier tenure periods and maximize wage increases in later tenure periods to provide incentive for high type workers. Subject to the constraint that firms' stage profits are nondecreasing after tenure period $T+1$, the contracts in which firms' stage profits are constant in all later tenure periods maximize wage increases in later tenure periods.

Notice that firms' stage profit π_t is increasing from tenure period 1 to $T-1$. By Lemma 3, π_T is greater than the weighted average of π_t $(1 \le t \le T-1)$, and π_t $(t \ge T+1)$ is equal to the weighted average of the stage profits from tenure period 1 to T. Therefore, $\pi_t \ge \pi_1$ for any t. As a result, $V_1 \ge \pi_1/(1-\delta) = y_1/(1-\delta) \ge p/(1-\delta)$. This implies that, if an optimal contract exists, firms have no incentive to deviate to the zero-wage contract (we can ignore requirement (iii) of the programming problem [PP]).

Lemma 3 indicates that optimal contracts are characterized by T $(T \ge 2)$ and w_T, the first tenure period in which wage is strictly positive and the wage in that tenure period. Once T and w_T are determined, w_{T+1} is determined by equation (11) and wages in later tenure periods are determined by the condition of constant stage profits. To ease exposition, we define

$$G(T, w_T) \equiv \sum_{t=T+1}^{\infty}(\delta\rho)^{t-1}(y_{t+1}-y_t) + (\delta\rho)^{T-2}w_T + (\delta\rho)^{T-1}(w_{T+1}-w_T), \tag{12}$$

$$\text{where } w_{T+1} = y_{T+1} - \frac{\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}y_t + (\delta\rho)^{T-1}\frac{\phi_1}{\phi_T}(y_T - w_T)}{\sum_{t=1}^{T}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}} \text{ by (11)}$$

Essentially, $G(T, w_T)$ is the LHS of (10) (or the discounted sum of wage increases) under the class of contracts specified in Lemma 3 with T and w_T, which measures the incentive provided to high type workers.

Lemma 4 *Fixing T, $G(T, w_T)$ is increasing in w_T. Define $g(T) \equiv \max_{w_T} G(T, w_T)$ subject to $w_T \leq w_{T+1}$. $g(T)$ is decreasing in T.*

Lemma 4 shows that, as w_T increases, the contract provides more incentive to high type workers. This is intuitive since an increase in w_T will lead to an increase in w_{T+1} by (11). Thus, the overall wage will increase more from tenure period $T-1$ to $T+1$, and a higher proportion of the wage increase occurs from tenure period $T-1$ to T. Both effects lead to more incentive provision. Essentially, $g(T)$ is the maximum incentive that could be provided given T. Lemma 4 shows that incentive provision is decreasing in T. Intuitively, an increase in T has two effects. First, it directly delays the wage increases. Second, it reduces the overall amount of wage increases. This is because it increases firms' stage profits in earlier tenure periods. To satisfy firms' no reneging conditions, it implies higher stage profits and lower wages in later tenure periods. Both effects lead to less incentive provision.

5.1 Existence of High-effort Equilibria

By Lemmas 3 and 4, the programming problem [PP] has a solution if and only if the maximum incentive provided with $T = 2$ is enough to motivate high type workers. More formally, the condition can be written as $g(2) \geq \widehat{c}$. Actually, the contract that provides the maximum incentive (corresponds to $g(2)$) is a constant-stage-profit contract (π_t is constant in t) in which the wage increases exactly match the increases in workers' expected productivity or the learning effect. More explicitly, the condition $g(2) \geq \widehat{c}$ can be written as

$$f(\phi_1) \equiv \sum_{j=1}^{\infty} (\delta\rho)^{j-1} (\phi_{j+1} - \phi_j) \geq \frac{\widehat{c}}{(1-p)}. \tag{13}$$

Note that ϕ_t (for all t) is a function of ϕ_1, the initial belief. Therefore, the left hand side of (13) is a function of ϕ_1, which we define as $f(\phi_1)$. Intuitively, $f(\phi_1)$ measures the maximum incentive provision given initial belief ϕ_1.

Lemma 5 *$f(0) = f(1) = 0$; $f(\phi_1)$ is strictly concave; $\frac{df}{d\phi_1}(0) > 0$ and $\frac{df}{d\phi_1}(1) < 0$.*

Lemma 5 is driven by the belief updating process. Essentially, $f(\phi_1)$ is the discounted sum of belief increases. When the initial belief is extreme (either 0 or 1), there is no belief updating. Thus $f(0) = f(1) = 0$. To understand that the discounted sum of belief increases is concave in initial belief, note that belief ϕ_t converges to 1 as t goes to infinity. If the initial belief ϕ_1 is high,

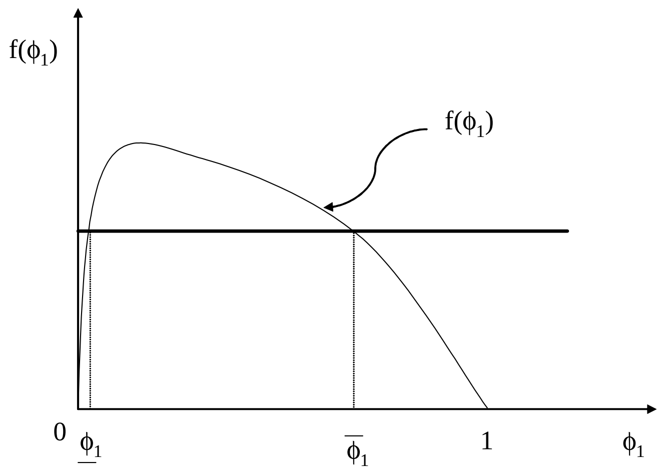

Figure 2: The Shape of $f(\phi_1)$

then the room for overall belief increase $(1-\phi_1)$ is small, which leads to a small discounted sum of belief increases. Conversely, when the initial belief ϕ_1 is small, although the room for overall belief increase is big, belief updating in earlier tenure periods will be slow. Due to discounting, the discounted sum of belief increases will be small as well.

Figure 2 plots a typical $f(\phi_1)$. Following Lemma 5, $f(\phi_1)$ is concave. Let $\phi_1^* \in (0,1)$ be the unique maximizer of $f(\phi_1)$, where ϕ_1^* is defined by $\frac{df}{d\phi_1}(\phi_1^*) = 0$. Thus, a necessary condition for condition (13) to hold is

$$f(\phi_1^*) \geq \frac{\widehat{c}}{(1-p)} = \frac{c}{\delta\rho(1-p)^2}. \tag{14}$$

Note that $f(\phi_1)$ is independent of c. Thus, (14) is satisfied if c is small enough. If (14) is satisfied, then there is an interval $[\underline{\phi}_1, \overline{\phi}_1]$ such that the programming problem [PP] has a solution for any $\phi_1 \in [\underline{\phi}_1, \overline{\phi}_1]$ (see figure 2), where $\underline{\phi}_1$ and $\overline{\phi}_1$ are the two roots of the equation $f(\phi_1) = \widehat{c}/(1-p)$. Also recall that, in the stationary state, ϕ_1 is a function of β with the following properties: $\phi_1(0) = 1$, $\phi_1(1) = 0$ and $d\phi_1/d\beta < 0$. It follows that the programming problem [PP] has a solution if and only if $\beta \in [\underline{\beta}, \overline{\beta}]$, where $\overline{\beta}$ is given by $\phi_1(\overline{\beta}) = \underline{\phi}_1$ and $\underline{\beta}$ is given by $\phi_1(\underline{\beta}) = \overline{\phi}_1$. The following proposition summarizes the previous analysis.

Proposition 3 *There exists $\underline{\beta}$ and $\overline{\beta}$, with $\underline{\beta} < \overline{\beta}$ and both interior to $[0,1]$, such that a high-effort equilibrium exists if only if (14) holds and $\beta \in [\underline{\beta}, \overline{\beta}]$.*

Proposition 3 implies that adverse selection helps alleviate moral hazard when firms are not able to commit to long-term contracts and all agents are able to change their partners freely in a market. This result is driven by the tension between incentive provision and contractual enforcement, as mentioned earlier. Incentive provision requires that the discounted sum of wage increases be big enough in order to motivate high type workers. However, contractual enforcement requires that the wage increases cannot exceed the increases of workers' expected output (the speed of learning), since, otherwise, longer-tenured workers are less profitable than new workers and firms will have incentives to renege. If there are no low type workers, then the learning effect is absent, and contractual enforcement requires that wage be constant. As a result, no incentive can be provided and high-effort equilibria cannot be sustained.[12] On the other hand, the presence of lower type workers can alleviate the tension between incentive provision and contractual enforcement by creating the learning effect. With the learning effect, workers' expected output is increasing with tenure, so an increasing wage contract can still satisfy the contractual enforcement constraint as long as wage increases more slowly than expected output does. Moreover, how fast wages can increase without violating the contractual enforcement constraint depends on the belief updating process. If the proportion of low type workers is too small, then the magnitude of belief updating is too small, thus not enough incentives can be provided. On the other end of spectrum, if the proportion of low type workers is too big, belief updates slowly in earlier tenure periods. Because of discounting, not enough incentives can be provided, either.

The assumption that both workers' past performance (in previous firms) and previous employment history are not observable (no record-keeping) is essential in sustaining high-effort equilibria. If a worker's past track-record were observable, then high-effort equilibria cannot be sustained, similar to the result in Mailath and Samuelson (2001). Intuitively, perfect record-keeping would destroy the learning effect in individual matches, which leads to a constant wage by the contract enforcement constraint, and, therefore, no incentive can be provided. Thus, in some sense, the absence of information flows among matches is beneficial in overcoming moral hazard in market settings.

Proposition 3 is derived under the assumption that there is no matching friction in markets. Matching friction, as shown by MacLeod and Malcomson (1998) and Yang (2008), also can alleviate moral hazard, since it generates a positive surplus in current employment relationships.[13] Broadly speaking, Proposition 3 can be understood as follows: to overcome moral hazard in markets, there must be some friction, and the friction can come from either matching friction or adverse selection.[14]

[12] This result is valid only if there is no matching friction in markets: workers and firms are of equal measure and there are no turnover costs. See the discussion two paragraphs later for more details.

[13] In MacLeod and Malcomson (1998), the matching friction comes from an unequal number of jobs and workers, while in Yang (2008), it comes from some exogenous turnover costs.

[14] In a repeated matching market with no contracts, Dutta (1993) and Ghosh and Ray (1996) show that high effort can still be *partially* sustained in equilibrium, even if there is no adverse selection or matching friction. One may wonder whether a similar result holds in our setting. Specifically, in the first N tenure periods, wages are zero, workers exert zero effort, and no endogenous separation occurs. After tenure period N, wages start to rise, high type workers exert high effort, and endogenous separation occurs. In the above strategy, the inefficiency endogenously created in early tenure periods might help to prevent parties from deviating.

However, if there is only matching friction, then stationary bonus contracts are optimal since the surplus of current employment relationships is independent of tenure. Given the fact that wage increases with tenure (thus not stationary) in reality, we believe that adverse selection plays a role in determining wage dynamics.

5.2 Optimal Contracts

Now we characterize the optimal contracts, assuming the programming problem [PP] has a solution ($g(2) \geq \widehat{c}$ holds).

Proposition 4 *Suppose (14) holds and* $\beta \in [\underline{\beta}, \overline{\beta}]$. *There is a unique optimal contract, which is quasi-monotonic and characterized by* T^* *and* $w^*_{T^*}$. *The unique* T^* *satisfies* $g(T^*) \geq \widehat{c}$ *and* $g(T^*+1) < \widehat{c}$, *and* $w^*_{T^*}$ *is determined by* $G(T^*, w^*_{T^*}) = \widehat{c}$. *In the optimal contract,* $w_t = 0$ *if* $t < T^*$, $w^*_{T^*+1}$ *is determined by (12), and for* $t > T^* + 1$ *the wage increases are equal to the increases in worker's expected output:* $w_{t+1} - w_t = y_{t+1} - y_t$.

The optimal contract is determined by three forces. First, to provide incentives to high type workers, the discounted sum of wage increases must be equal to a given level. Second, to reduce informational rent to low type workers, firms try to backload wages as much as possible since low type workers are more likely to be in earlier tenure periods. Third, firms' ability to backload wages is limited by firms' no-reneging conditions: senior workers have to be more profitable than new workers. The last two forces pin down the form of optimal contracts: firms' stage profits are constant in later tenure periods and wages in early tenure periods are low and constant (zero). By Lemma 4, within this form of contracts, backloading wages more (delaying the first tenure period in which wage is strictly positive) leads to less incentive provision. In the optimal contract, wages are backloaded as much as possible subject to (just) enough incentive being provided to high type workers.

To get a better feel for the contract dynamics, Figure 3 illustrates the evolution of firms' stage profits π_t and firms' discounted expected payoff V_t under the optimal contract. In this example, wage starts to increase in tenure period 3 ($T^* = 3$). Firms' stage profits first increase, then reach their maximum in tenure period 2, decrease in tenure periods 3 and 4, and remain constant in all later tenure period. Firms' discounted payoff exhibits the same pattern. Moreover, for $t \geq 4$, $V_t = V_1$.

By Proposition 4, the wage dynamics in the optimal contract exhibits two salient features. First, wage is low and remains constant in earlier tenure periods. Second, when wage starts to increase in later tenure periods, it is intimately related to the learning effect: the wage increase in each tenure period (after tenure period T^*) is exactly equal to the increase in a worker's expected output. This is because the stage profit is constant after wage starts to increase.

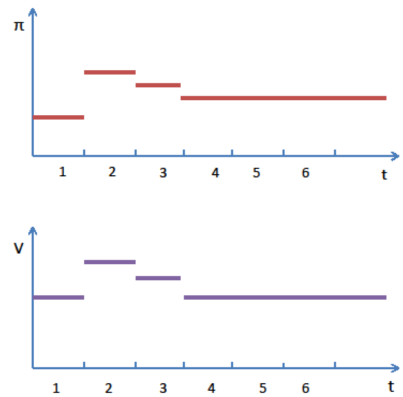

Figure 3: The evolution of firms' stage profits and discounted payoff

The wage dynamics exhibited by the optimal contract in our model are different from those in Jovanovic (1979) in two aspects. First, in his model, wage in each tenure period is exactly equal to a worker's expected productivity. In contrast, in our model, wages are low and constant in earlier tenure periods, and only the *wage increase* in later tenure periods is equal to the increase in a worker's expected productivity. Second, in his model, the fact that wage always matches a worker's expected productivity is due to market competition. In contrast, in our model, workers' performance is not observed by the market; yet, in the optimal contract, the wage increases are intrinsically tied to workers' performance. This implies that, even without market competition, wages being tied to workers' expected productivities can be generated by *internal wage dynamics*. In our model, it is the interaction between incentive provision and contractual enforcement that leads to the internal wage dynamics being tied to the learning effect. Note that both incentive provision and contractual enforcement are indispensable. Suppose there is no moral hazard, then wage need not increase to provide incentives. On the other hand, suppose contractual enforcement is not an issue (say firms are able to commit to long-term contracts), then the wage increases can be arbitrary, as firms can backload wages as much as possible.[15]

[15] The output being not verifiable is a key assumption that leads to a lack of intertemporal commitment. Typically, the length of an employment relationship (tenure period t) is verifiable. If output were verifiable, then the following contract can be enforced by the court: the firm will continue the relationship if and only if the output is 1, and if the relationship continues, the wage follows $\{w_t\}$. Now firms cannot renege and they can backload wages as much as they want. With output not being verifiable and firing being necessary for workers producing zero output, firms always have flexibility in terminating a relationship. This means that firms cannot commit to long-term contracts

Limited liability is essential in deriving the optimal contract. Basically, limited liability ensures that low type workers will get positive rents by mimicking high type workers. This creates an incentive for firms to backload wages in order to reduce low type workers' informational rent. Without limited liability, firms can change the wage profile (change wages in all tenure periods by a constant) such that low type workers always get zero rent. As a result, firms would have no incentive to backload wages since low type workers will get zero rent anyway, although they still need to backload wages in order to provide incentive for high type workers. This implies that there are many optimal contracts and the wage dynamics in optimal contracts can be of various form.[16]

To better understand the bite of the "relational" aspect on the optimal contracts, here we briefly discuss the form of the optimal long-term contracts. Since firms' reneging is not an issue under long-term contracts, firms will try to backload wage as much as possible in order to reduce the information rent for low type workers. Therefore, the form of the optimal long-term contracts is as follows: wage starts at zero and remains at zero for many tenure periods, then at some tenure period there is a significant jump of wage and it remains constant in all later tenure periods. Compared to the optimal long-term contracts, we can clearly see that the "relational" aspect of contracts limits firms' ability to backload wages, which ties the wage increases in the optimal (relational) contract to the learning effect.

So far we have focused on contracts that maximize firms' expected profit. Contracts that maximize firms' and high type workers' joint surplus would have the same qualitative feature as the contracts maximizing firms' profit. This is because both require that the payoff to low type workers be minimized, so wages should be backloaded as much as possible. Of course, as high type workers' share of the surplus increases, low type workers' payoff will increase as well, as they can always mimic high type workers with some success.

6 Separating Contracts

Given that there are two types of workers, theoretically speaking, firms could offer separating contracts: firms offers two contracts and let workers self-select in tenure period 1. Specifically, in the contract designed for L type workers, a fixed wage w_L is offered in tenure period 1, and the worker is fired after tenure period 1 regardless of the output. In the contract designed for H type

even if tenure period t is verifiable.

[16] Specifically, without limited liability for optimal contracts, the binding constraints are high type workers' no-shirking condition, low type workers' individual rationality (IR) condition, and firms' no-reneging conditions. Essentaily, the binding no-shirking condition of high type workers dictates the wage increases while the binding IR condition of low type workers dictates the starting wage given any profile of wage increases. There could be many contracts with different combinations of starting wages and wage increase patterns satisfying those three constraints, and hence all of them are optimal. For example, in one optimal contract, the majority of wage increases could occur in the early tenure periods while, in another one, the majority of wage increases could occur in later tenure periods (the discounted sum of wage increases are the same under two contracts due to the binding no-shirking condition of high type workers). The binding IR condition of low type workers requires that the starting wage is lower in the first optimal contract than in the second optimal contract.

workers, the payment plan evolves according to $\{w_t^s\}$. In short, we denote a separating contract as $(w_L, \{w_t^s\})$. In the rest of this section, we just provide the main results regarding separating contracts. The detailed analysis can be found in an online appendix.

Compared to pooling contracts, separating contracts lead to several differences. First, with separating contracts firms learn the type of new workers in the first tenure period, while under pooling contracts the learning is gradual. Second, with separating contracts there are two additional self-selection constraints: L type workers have no incentive to choose the contract for type H, and H type workers have no incentive to choose the contract for type L. Without loss of generality, w_L should be set such that type L workers are indifferent between choosing the contract for type H and choosing the contract for type L. That is, w_L equals the average per-period payoff if an L type chooses the H contract $\{w_t^s\}$. It turns out that the self-selection constraint for type H is redundant. Intuitively, if an H type worker chooses contract w_L, he gets the same payoff when he chooses contract $\{w_t^s\}$ and shirks in every period. Therefore, no-shirking conditions imply that H type workers have no incentive to choose the L contract.

The self-selection constraint for L types deserves further comment. In typical repeated adverse selection models (e.g., Laffont and Tirole, 1990), inducing separation in the first period is very costly, as the discounted sum of informational rents in all future periods has to be paid in the first period. Translating into our setting, w_L would have been $\sum_{t=1}^{\infty}(\delta\rho p)^{t-1}w_t^s$, the discounted informational rents in a relationship. However, in our model w_L is less than the discounted informational rents in a relationship. The difference is that, in repeated adverse selection models, there is only a single relationship and thus an agent gets zero rent after revealing his type. In contrast, in our model this is not the case: after leaving the current relationship, next period an L type worker can match with another firm and get informational rents as well. Therefore, there is an opportunity cost for an L type worker to mimic an H type in the current relationship. As a result, to induce an L type worker to reveal his type, a firm does not need to pay the discounted sum of informational rents in the current relationship. In other words, the cost of separating is relatively low. Define the cost of separating as w_L minus the per-period wage that firms pay on average to an L type worker who always chooses the H contract. Note that the latter equals $(1-\delta\rho p)\sum_{t=1}^{\infty}(\delta\rho p)^{t-1}w_t$. Therefore, in the current setup of the model, the cost of separating is zero.

With separating contracts, high-effort equilibria exist if and only if there are enough L type workers.[17] The wage increases of $\{w_t^s\}$ has to be big enough to motivate H type workers. To prevent firms from reneging, there must be enough punishment for reneging. This punishment comes from the scarcity of H type workers who generate higher profits for firms: after reneging, firms must match with new workers who might be L type workers. The more L type workers, the lower the probability to match with an H type worker in the unmatched pool, hence the bigger the

[17] The formal condition is that $\beta \in [\widehat{\beta}, 1]$, $\widehat{\beta} \in (0, 1)$.

punishment for reneging. Recall Proposition 3. Self-enforcing pooling contracts exist if and only if the proportion of L type workers is not too low or too high. The difference comes from the fact that, with pooling contracts, the wage increases cannot exceed the speed of learning. When the proportion of L type workers is too high, the belief updating will be very slow initially, and due to discounting, not enough incentives can be provided to H type workers.

If high-effort equilibria exist, then the optimal separating contract (which maximizes firms' expected profit) is unique. Specifically, the optimal separating contract has the following form: the wage is zero in early tenure periods, and then it increases for two tenure periods and stays constant in later tenure periods. The forces that determine the optimal separating contract are similar to those that govern the optimal pooling contract. To provide incentives to high type workers, the discounted sum of wage increases must be big enough. To reduce informational rent to low type workers, firms try to backload wages as much as possible. However, firms' ability to backload wages is limited by firms' no-reneging conditions. The last two forces lead to constant stage profits in later tenure periods and constant (zero) wage in early tenure periods.

The wage dynamics in the optimal separating contract and the optimal pooling contract share a similar feature: wage is low and remains constant in earlier tenure periods. The difference is that, in the optimal separating contract, wage increases at most in two tenure periods, and then wage remains constant afterwards. This difference comes from the fact that learning is completed in the first tenure period under separating contracts. Thus, constant stage profits in later tenure periods imply constant wage.

Compared to pooling contracts, high-effort equilibria exist for a wider range of parameter values under separating contracts. The intuition for this result is as follows. Under separating contracts, since learning is completed in the first tenure period, subject to the no-reneging conditions, the maximum amount of wage increase can occur in the second tenure period. On the other hand, since under pooling contracts learning occurs gradually, the same amount of wage increase has to be spread over many tenure periods. Due to discounting, less incentive is provided to H type workers with pooling contracts.

If higher-effort equilibria exist under both contracts, it can be shown that the optimal separating contract yields a higher profit for firms than the optimal pooling contract does. The intuition for the comparison is as follows. Under both types of contracts, firms' ability to backload wages are more or less the same. Under pooling contracts, firms' ability to backload wages is dictated by the gradual increase of beliefs about workers as tenure period increases. Under separating contracts, though learning is completed in tenure period 1, firms are able to backload wages since, in tenure period 1, workers are very likely to be of low type. Comparing separating contracts and pooling contracts, there is an additional effect that favors separating contracts. With separating contracts, a firm is able to learn the type of a new worker immediately. In contrast, with pooling contracts it takes a longer time for a firm to learn a worker's type. Thus, with the same initial beliefs, on average it takes a shorter time for a firm to match with an H type worker with separating contracts. This fast screening effect favors separating contracts.

Although the optimal separating contract yields a higher profit for firms, we believe that pooling contracts are more relevant: in the real world, we seldom see firms offer multiple contracts to common workers and let them self-select. For example, Bewley (1999) found that firms intentionally avoid paying workers assigned to similar tasks differentiated wages. This suggests that pooling contracts are more likely to be the social norm.[18]

Recall that there is no cost of separating in our current setting. However, in the real world the following factors (outside the model) tend to give rise to the cost of separating and favor pooling contracts. First, offering different wages might dampen workers' overall morale, as shown in Fang and Moscarini (2005). Specifically, they show that, when workers' effort and ability are complements in production, firms do not want to offer different wage contracts to workers. Second, it might be the case that with some positive probability previous employment history is observed by new employers. Thus, always selecting the L contract and getting fired immediately will cause some new employers to offer zero wage instead, which makes L type workers reluctant to choose the L contract. Finally, turnover costs in the real world will lead to a positive cost of separating. The turnover costs could be physical costs involved in changing jobs or due to the presence of unemployment. Intuitively, when the turnover costs are present, L types have incentives to pool with H types in the current match to avoid separation, since separation now will lead to a lower continuation payoff. As a result, to induce immediate type revelation, the firm has to pay a higher w_L to compensate for separation.[19] As the turnover costs increases, in order to induce immediate type revelation the payment to an L type becomes closer to the discounted sum of informational

[18] Note that the (relational) contract serves as some sort of social norm, which pins down agents' beliefs on and off the equilibrium path.

[19] To illustrate the idea, consider the case of positive unempolyment. Suppose the measure of workers is still 1, but the measure of firms is $\alpha \in (0,1)$. That is, workers are on the long side of the market. Now, workers in the unmatched pool may not get a match. Let $\gamma \in (0,1)$ be the probability that a worker in the unmatched pool matches with a firm. Obviously, γ is increasing in α. Suppose firms offer both types a pooling contract $\{w_t\}$. Define U_u^L as a L type's payoff when he is in the unmatched pool, and U_1^L as a L type's payoff when he is just matched with a firm and always chooses the H contract.

$$\begin{aligned} U_u^L &= \gamma U_1^L + (1-\gamma)\delta U_u^L; \\ U_1^L &= \sum_{t=1}^{\infty} (\delta\rho p)^{t-1} w_t + \frac{\delta(1-\rho p)}{1-\delta\rho p} U_u^L. \end{aligned}$$

Now to induce immediate type revelation,

$$w_L = U_1^L - \delta U_u^L = \frac{1-\delta\rho p}{1-(1-\gamma)\delta\rho p} \sum_{t=1}^{\infty} (\delta\rho p)^{t-1} w_t > (1-\delta\rho p) \sum_{t=1}^{\infty} (\delta\rho p)^{t-1} w_t,$$

where $(1-\delta\rho p)\sum_{t=1}^{\infty}(\delta\rho p)^{t-1} w_t$ is the average per-period wage that firms pay to L type workers in the pooling contract. Thus, the separating cost now is positive. It is easy to see that, as γ (α) decreases, w_L increases and thus the cost of separating increases.

rents in the current relationship.[20]

7 Conclusions

This paper studies nonstationary relational contracts driven by the presence of adverse selection. The internal wage dynamics are pinned down by the tension between incentive provision and contractual enforcement. The paper contributes to the understanding of how contractual enforcement restricts firms' ability in offering long-term contracts in nonstationary environments. Moreover, the paper shows that, when contractual enforcement is an issue and agents are free to change partners in markets, adverse selection can alleviate moral hazard.

Although our model is framed in a labor market setting, it can be applied to broader settings. In fact, it applies to markets in which both moral hazard and adverse selection exist and contractual enforcement is an issue. The internal wage dynamics derived in our model can be generalized as *internal contract dynamics*. Two relevant examples are lending markets and buyer-seller relationships. In the context of lending markets,[21] our model implies that, as a lending relationship continues, the contractual terms should become more favorable to the borrower, who has the moral hazard problem. This is consistent with the phenomenon of relationship lending: borrowers with longer relationships with a bank pay lower interest rates and are less likely to pledge collateral (Berger and Udell, 1995; Bodenhorn, 2003).

Appendix: Proofs

Proof of Lemma 2.

Proof. (i) The no-shirking conditions can be reduced to (10). First, we show that under $\{w_t\}$, $U_t \geq U_2$ for any $t \geq 3$. By (3),

$$U_t = \frac{\delta(1-\rho)}{1-\delta\rho}U_1 - \frac{1}{1-\delta\rho}c + \sum_{l=t}^{\infty}(\delta\rho)^{l-t}w_l. \tag{15}$$

By (15), for $t \geq 3$

$$U_t - U_2 = \sum_{j=t}^{\infty}(\delta\rho)^{j-t}w_j - \sum_{j=2}^{\infty}(\delta\rho)^{j-2}w_j = \sum_{j=0}^{\infty}(\delta\rho)^{j}(w_{t+j} - w_{2+j}) \geq 0,$$

since w_t is nondecreasing in t. Therefore, the no-shirking conditions boil down to $U_2 - U_1 \geq \widehat{c}$, which can be explicitly written as (10).

[20] Introducing turnover costs would not change the optimal pooling contract qualitatively, though low type workers would get less informational rent.

[21] Specifically, consider a lending market with two types of borrowers (firms): high type and low type. High type firms can choose to implement one project from two available projects: one bad project which is more risky, and one good project with a safe and higher expected return. Low type firms only have access to the bad project.

(ii) If $\pi_{t+1} \geq \pi_t$ for any $t \geq T$, then $V_T \geq V_1$ implies firms' no-reneging conditions (9). To prove this claim, we first show that $V_t \geq V_1$ for any $t > T$. Suppose $V_{T+1} < V_1$. Then, combining with $V_T = \pi_T + \delta[\rho\frac{\phi_T}{\phi_{T+1}}V_{T+1} + (1-\rho\frac{\phi_T}{\phi_{T+1}})V_1] \geq V_1$, we have $\pi_T > (1-\delta)V_1$. But

$$\begin{aligned} V_{T+1} &= \sum_{j=T+1}^{\infty}(\delta\rho)^{j-(T+1)}\frac{\phi_{T+1}}{\phi_j}\pi_j + \delta[1-(1-\delta)\sum_{j=T+1}^{\infty}(\delta\rho)^{j-T}\frac{\phi_{T+1}}{\phi_{j+1}}]V_1 \\ &\geq \sum_{j=T+1}^{\infty}(\delta\rho)^{j-(T+1)}\frac{\phi_{T+1}}{\phi_j}(1-\delta)V_1 + \delta[1-(1-\delta)\sum_{j=T+1}^{\infty}(\delta\rho)^{j-T}\frac{\phi_{T+1}}{\phi_{j+1}}]V_1 = V_1, \end{aligned}$$

where the inequality follows $\pi_{t+1} \geq \pi_t$ for any $t \geq T$ and $\pi_T > (1-\delta)V_1$; a contradiction. Therefore, $V_{T+1} \geq V_1$. By similar arguments, we can recursively show that $V_t \geq V_1$ for any $t > T$.

Next we show that $V_t \geq V_1$ for any $t < T$ (this step is necessary only if $T > 2$). Suppose $V_{T-1} < V_1$. Then $V_{T-1} = \pi_{T-1} + \delta[\rho\frac{\phi_{T-1}}{\phi_T}V_T + (1-\rho\frac{\phi_{T-1}}{\phi_T})V_1] < V_1$ and $V_T \geq V_1$ implies that $\pi_{T-1} < (1-\delta)V_1$. Since π_t is increasing in the domain from 1 to T, We have $V_{T-2} = \pi_{T-2} + \delta[\rho\frac{\phi_{T-2}}{\phi_{T-1}}V_{T-1} + (1-\rho\frac{\phi_{T-2}}{\phi_{T-1}})V_1] < V_1$. Applying this argument recursively, eventually we have, $V_1 = \pi_1 + \delta[\rho\frac{\phi_1}{\phi_2}V_2 + (1-\rho\frac{\phi_1}{\phi_2})V_1] < V_1$, a contradiction. Therefore, $V_{T-1} \geq V_1$. By similar arguments, we can show that $V_t \geq V_1$ for any $t < T$.

(3) If $\pi_{t+1} \geq \pi_t$ for any $t > T$ and $\pi_T > \pi_{T+1}$, then $V_{T+1} \geq V_1$ implies that $V_t \geq V_1$ for any t. Recall that this type of quasi-monotonic contract must have $\pi_T \geq (1-\delta)V_1$. Combining with $V_{T+1} \geq V_1$, we have $V_T \geq V_1$. The rest of the proof is the same as the proof in (2). ■

Proof of Proposition 2.

Proof. (i) Suppose a quasi-monotonic contract $\{w_t\}$ satisfies the no-shirking and no-reneging conditions. Moreover, $U_2 - U_1 > \widehat{c}$. Let j be a tenure period such that $w_{j+1} > w_j$ (such a j must exists, otherwise (10) is violated, and $j \geq T$). The idea is to find another self-enforcing quasi-monotonic contract which yields a strictly larger V_1. Specifically, construct another contract $\{w_t'\}$ as follows: $w_t' = w_t$ for any $t \leq j$, $w_t' = w_t - \varepsilon$ for any $t > j$, where $\varepsilon > 0$ is very small. By construction, $\{w_t'\}$ is also quasi-monotonic. By the construction, it is easy to see that $V_j' > V_j$ and $V_1' > V_1$.

Now what is left to be shown is that $\{w_t'\}$ is self-enforcing. $\{w_t'\}$ clearly satisfies the no-shirking condition (10). To see this, note that compared with $\{w_t\}$, under $\{w_t'\}$ only the wage increase from j to $j+1$ is reduced by ε. From (10), we can see that that $U_2 - U_1 > \widehat{c}$ implies that $U_2' - U_1' \geq \widehat{c}$. The next step is to show that $\{w_t'\}$ satisfies the no-reneging conditions. Since $\{w_t'\}$ is quasi-monotonic, we only need to show $V_T' \geq V_1'$. Note that V_1' increases because V_T' increases. Given that

$$\{1-\delta\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}[1-\rho\frac{\phi_t}{\phi_{t+1}}]\}V_1 = \sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}\pi_t + (\delta\rho)^{T-1}\frac{\phi_1}{\phi_T}V_T$$

and the same relationship holds between V_T' and V_1', we have $V_1' - V_1 < V_T' - V_T$. Since $V_T \geq V_1$ ($\{w_t\}$ satisfies the no-reneging conditions), we must have $V_T' \geq V_1'$. This proves part (i).

Part (ii) is directly implied by part (i) and Proposition 1. ■

Proof of Lemma 3.

Proof. Part (i). Suppose there is a self-enforcing quasi-monotonic contract $\{w_t\}$ in which $\pi_{T+2} > \pi_{T+1}$. We want to show that there is another self-enforcing quasi-monotonic contract $\{w'_t\}$ which yields a higher expected profit for firms (a similar argument can be applied for later tenure periods). From the original contract $\{w_t\}$, which satisfies (10) and (ICF), we construct another $\{w'_t\}$ as follows: increase w_t by ε (ε is small) for any $t \geq T+2$, and decrease w_{T+1} by $\Delta = \sum_{t=T+2}^{\infty} (\delta\rho)^{t-(T+1)} \frac{\phi_{T+1}}{\phi_t}\varepsilon$. Note that by construction $\{w'_t\}$ is also quasi-monotonic. Moreover, $V'_{T+1} = V_{T+1}$, $V'_T = V_T$ and $V'_1 = V_1$. Therefore, the no-reneging conditions (ICF) hold under $\{w'_t\}$. Now consider the no-shirking condition (10). The change of the LHS of (10) is

$$\begin{aligned} & (\delta\rho)^{T-1}[(1-\delta\rho)(w'_{T+1} - w_{T+1}) + \delta\rho(w'_{T+2} - w_{T+2})] \\ = \; & (\delta\rho)^{T-1}\varepsilon[\delta\rho - (1-\delta\rho)\sum_{t=T+2}^{\infty} (\delta\rho)^{t-(T+1)} \frac{\phi_{T+1}}{\phi_t}] > (\delta\rho)^{T-1}\varepsilon[\delta\rho - (1-\delta\rho)\frac{\delta\rho}{1-\delta\rho}] = 0. \end{aligned}$$

Therefore, under $\{w'_t\}$ (10) is satisfied and not binding. By part (i) of Proposition 2, both $\{w'_t\}$ and the original contract $\{w_t\}$ cannot be optimal. Therefore, we must have $\pi_{T+2} = \pi_{T+1}$ in optimal contracts.

Part (ii). Suppose $V_{T+1} > V_1$ in the original quasi-monotonic contract $\{w_t\}$, which satisfies (10) and (ICF). We construct another contract $\{w'_t\}$ as follows: increase w_t by ε for all $t \geq T+1$ and reduce w_T by $\Delta = \sum_{t=T+1}^{\infty} (\delta\rho)^{t-T} \frac{\phi_1}{\phi_t}\varepsilon$. The new contract $\{w'_t\}$ is still quasi-monotonic. $V'_{T+1} < V_{T+1}$ and $V'_1 = V_1$. But for ε small enough $V'_{T+1} \geq V'_1$ still holds, since $V_{T+1} > V_1$. Therefore, the no-reneging conditions (ICF) are satisfied under $\{w'_t\}$. As in the proof of part (i), it can be verified that (10) is satisfied and not binding under $\{w'_t\}$. But a nonbinding (10) implies that both $\{w'_t\}$ and the original contract $\{w_t\}$ are not optimal. Therefore, we must have $V_{T+1} = V_1$ in optimal contracts.

In optimal contracts, given that π_t is constant after tenure period $T+1$, and $V_{T+1} = V_1$, we must have $\pi_{T+1} = (1-\delta)V_1$. Moreover, V_t is constant after tenure period $T+1$ as well. Writing V_1 recursively and using $V_{T+1} = V_1$, we have

$$(1-\delta)V_1 \sum_{t=1}^{T} (\delta\rho)^{t-1} \frac{\phi_1}{\phi_t} = \sum_{t=1}^{T} (\delta\rho)^{t-1} \frac{\phi_1}{\phi_t} \pi_t,$$

which gives rise to (11). ■

Proof of Lemma 4.

Proof. Inspecting (12), we see that w_{T+1} is increasing in w_T. Since $G(T, w_T)$ is increasing in both w_T and w_{T+1}, $G(T, w_T)$ is increasing in w_T. However, the restriction of $\pi_T \geq (1-\delta)V_1$ places an upper bound on w_T. Substituting in this upper bound, we have

$$g(T)=\sum_{t=T}^{\infty}(\delta\rho)^{t-1}(\frac{\phi_{t+1}}{\phi_{t+2}}-\frac{\phi_t}{\phi_{t+1}})+(\delta\rho)^{T-2}[\frac{\phi_T}{\phi_{T+1}}-\frac{\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_{t+1}}}{\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}}]. \quad (16)$$

Inspecting (16), we see that $g(T)$ is decreasing in T. ■

Proof of Lemma 5.

Proof. It is easy to verify that $f(0)=f(1)=0$, since $\phi_t=0$ (for all t) if $\phi_1=0$ and $\phi_t=1$ (for all t) if $\phi_1=1$. Expand $f(\phi_1)$ and take the derivative with respect to ϕ_1:

$$\begin{aligned}\frac{df}{d\phi_1} &= -1+(1-\delta\rho)\sum_{t=1}^{\infty}(\delta\rho)^{t-1}\frac{1}{\phi_1+p^t(1-\phi_1)} \\ &\quad +\phi_1(1-\delta\rho)\sum_{t=1}^{\infty}(\delta\rho)^{t-1}\frac{p^t-1}{[\phi_1+p^t(1-\phi_1)]^2}.\end{aligned}$$

Note that

$$\begin{aligned}\frac{df}{d\phi_1}(0) &= (1-\delta\rho)\sum_{t=1}^{\infty}(\delta\rho)^{t-1}(\frac{1}{p^t}-1)>0; \\ \frac{df}{d\phi_1}(1) &= (1-\delta\rho)\sum_{t=1}^{\infty}(\delta\rho)^{t-1}(p^t-1)<0.\end{aligned}$$

Take the second derivative,

$$\frac{d^2f}{d\phi_1^2}=2(1-\delta\rho)\sum_{t=1}^{\infty}(\delta\rho)^{t-1}\{\frac{p^t-1}{[\phi_1+p^t(1-\phi_1)]^2}+\phi_1\frac{(p^t-1)^2}{[\phi_1+p^t(1-\phi_1)]^3}\}.$$

Note that the term in the bracket is

$$\frac{(p^t-1)}{[\phi_1+p^t(1-\phi_1)]^2}+\phi_1\frac{(p^t-1)^2}{[\phi_1+p^t(1-\phi_1)]^3}=\frac{-(1-p^t)p^t}{[\phi_1+p^t(1-\phi_1)]^3}<0$$

Therefore, $\frac{d^2f}{d\phi_1^2}<0$ or $f(\phi_1)$ is strictly concave. ■

Proof of Proposition 4.

Proof. By Proposition 2 and Lemma 3, optimal contracts must be quasi-monotonic and are characterized by T and w_T. Given T and w_T, w_{T+1} can be computed according to (12). For all $t<T$, $w_t=0$, and for $t>T+1$, $w_t=w_{T+1}+(y_t-y_{T+1})$ by the constant-stage-profit requirement. The binding (10) pins down optimal contracts: $G(T,w_T)=\widehat{c}$. We first determine T^*. By Lemma 4, $g(T)$ is decreasing in T. Moreover, $\lim_{T\to\infty}g(T)=0$ and $g(2)\geq\widehat{c}$. Therefore, there is a unique T^* such that $g(T^*)\geq\widehat{c}$ and $g(T^*+1)<\widehat{c}$. Given T^*, there is a unique $w^*_{T^*}\in(0,\frac{\phi_T}{\phi_{T+1}}-\frac{\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_{t+1}}}{\sum_{t=1}^{T-1}(\delta\rho)^{t-1}\frac{\phi_1}{\phi_t}})$ such that $G(T^*,w^*_{T^*})=\widehat{c}$. Therefore, the optimal contract is unique. ■

References

[1] Baker, George, Robert Gibbons and Kevin Murphy. "Subjective Performance Measures in Optimal Incentive Contracts," *Quarterly Journal of Economics*, 1994, Nov., 1125-56.

[2] Becker, Gary. "Investment in Human Capital: A Theoretical Analysis." *Journal of Political Economy*, 1962, Oct., 9-49.

[3] Becker, Gary. *Human Capital*. Second edition, 1975, Chicago, University of Chicago Press.

[4] Berger, Allen and Gregory Udell. "Relationship Lending and Lines of Credit in Small Firm Finance." *Journal of Business*, 1995, Jun., 351-81.

[5] Bernhardt, Dan. "Strategic Promotion and Compensation." *Review of Economic Studies*, 1995, Apr., 315-39.

[6] Bewley, T.F. *Why Wages Don't Fall During a Recession*, 1999. Harvard University Press, Cambridge, MA.

[7] Board, S. and M. Meyer-ter-Vehn. "Relational Contracts in Competitive Labor Markets," 2011, working paper.

[8] Bodenhorn, Howard. "Short-Term Loans and Long-Term Relationships: Relationship Lending in Early American." *Journal of Money, Credit and Banking*, 2003, Aug., 485-505.

[9] Bull, Clive. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics*, 1987, Feb., 147-59.

[10] Burdett, Ken and Melvyn Coles. "Equilibrium Wage-Tenure Contracts." *Econometrica*, 2003, Sept., 1377-1404.

[11] Chassang, S. "Building Routines: Learning, Cooperation, and the Dynamics of Incomplete Relational Contracts," *American Economic Review,* 2010, 448-465.

[12] Dutta, "Building Trust", 1993, Mimeo. London School of Economics.

[13] Halac, M. "Relational Contracts and the Value of Relationships," *American Economic Review*, forthcoming.

[14] Fang, H. and G. Moscarini. "Morale Hazard," *Journal of Monetary Economics*, 2005, 749-777.

[15] Farber, Henry and Robert Gibbons. "Learning and Wage Dynamics." *Quarterly Journal of Economics*, 1996, Nov., 1007-47.

[16] Felli, Leonardo and Chris Harris. "Learning, Wage Dynamics, and Firm-Specific Human Capital." *Journal of Political Economy*, 1996, Nov., 838-68.

[17] Fong, Yuk-fai, and Jin Li. "Relational Contracts, Efficiency Wages, and Employment Dynamics," 2010, mimeo.

[18] Ghosh, P. and Ray, D. "Cooperation in Community Interaction without Information Flows", *Review of Economic Studies*, 1996, 63(3), 491-519.

[19] Greenwald, Bruce. "Adverse Selection in the Labor Market." *Review of Economic Studies*, 1986, Jun., 325-47.

[20] Harris, Milton and Bengt Holmstrom. "A Theory of Wage Dynamics." *Review of Economic Studies*, 1982, Jul., 315-33.

[21] Hashimoto, Masanori. "Firm-Specific Human Capital as a Shared Investment." *American Economic Review*, 1981, Jun., 475-82.

[22] Holmstrom, Bengt. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies*, 1999, Jan., 169-82.

[23] Jovanovic, Boyan. "Job Matching and the Theory of Turnover." *Journal of Political Economy*, 1979, Oct., 972-90.

[24] Kranton, R. "The Formation of Cooperative Relationships," *Journal of Law, Economics, & Organization*, 1996, 214-233.

[25] Laffont, J.J. and J. Tirole. "Adverse Selection and Renegotiation in Procurement," *Review of Economic Studies*, 1990, 597-626.

[26] Lazear, Edward. "Why Is There Mandatory Retirement." *Journal of Political Economy*, 1979, Dec., 1261-84.

[27] Levin, Jonathan. "Relational Incentive Contracts." *American Economic Review*, 2003, Jun., 835-847.

[28] MacLeod, Bentley and James Malcomson. "Reputation and Hierarchy in Dynamic Models of Employment," *Journal of Political Economy*, 1988, 96(4), 832-854.

[29] MacLeod, Bentley and James Malcomson. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica*, 1989, Mar., 447-80.

[30] MacLeod, Bentley and James Malcomson. "Motivation and Markets." *American Economic Review*, 1998, Jun., 388-411.

[31] Mailath, George and Larry Samuelson. “Who Wants a Good Reputation.” *Review of Economic Studies*, Apr. 2001, 415-441.

[32] Mincer, Jacob. *Schooling, Experience, and Earnings*. 1974, New York, NBER.

[33] Rob, Rafael and Huanxing Yang. “Long-Term Relationships as Safeguards,” *Economic Theory*, 43(2), 2010, 143-166.

[34] Thomas, J. and T. Worrall. “Dynamic Relational Contracts with Credit Constraints,” 2010, working paper.

[35] Waldman, Michael. “Job Assignments, Signaling and Efficiency.” *Rand Journal of Economics*, 1984, Summer, 255-67.

[36] Yang, Huanxing. “Efficiency Wages and Subjective Performance Pay.” *Economic Inquiry*, 46(2), 2008, 179-196.