Efficiency Wages and Performance Pay

Huanxing Yang‡
The Ohio State University
January 5, 2005

Abstract

This paper studies contract selection between efficiency wages and subjective performance pay to motivate workers in a market setting. The optimal wage contract (which may be a combination of efficiency wages and subjective performance pay) is determined by turnover costs in labor markets. Specifically, labor markets with higher turnover costs use more subjective performance pay and less efficiency wages. As a result, in these markets the total wage payment is lower and equilibrium employment level is higher. Surprisingly, under certain conditions an increase in turnover costs leads to higher social welfare. In an extended model incorporating workers’ search costs, we show that wages are procyclical in booms, and are either rigid or counter-cyclical during recessions. The predictions of the model are consistent with some empirical evidence.

JEL Classification: D82, J33, J41, J63.

‡Department of Economics, The Ohio State University, 405 Arps Hall, 1945 N. High Street, Columbus, OH 43210.
Email: yang.1041@osu.edu.

†This is a revised version of the second chapter of my dissertation presented to the University of Pennsylvania. I wish to thank Ken Burdett, Bill Dupor, Jan Eeckhout, Paul Evens, Rafael Rob, and seminar participants at the Midwest Theory Conference 2003 at Indiana University at Bloomington for helpful comments and discussions. I am truly indebted to Steven Matthews for his invaluable advice, support and guidance. The usual disclaimers apply.
1 Introduction

How to design incentive schemes to motivate workers is an important topic both in micro and macro economics. The shirking models of efficiency wages, (like Shapiro and Stiglitz, 1984) establish that firms need to pay a wage premium (efficiency wages) to motivate workers, and unemployment serves as a punishment device. However, one shortcoming of these models is that performance pay plays no role. One justification for their ignorance of performance pay is that individual performance is difficult to be measured in an objective way, thus the use of legally enforceable performance pay is practically limited. Nevertheless, if workers’ performance is observable and employment relationships are repeated, firms can use implicit bonuses or relational contracts (Bull, 1987; MacLeod and Malcomson, 1989; Baker, Gibbons and Murphy, 1994), based on workers’ subjectively assessed performance, to motivate workers. Actually, subjective performance pay are widely observed in practice. Since subjective performance pay cannot be legally enforced, it has to be self-enforcing. This requirement may restrict the use of subjective performance pay in practice.

Given that both efficiency wages and subjective performance pay are able to motivate workers, what is the optimal wage contract from firms’ perspective? Will different labor markets (occupations) use different forms of wage contracts? What are the impacts of different forms of wage contracts on unemployment and social welfare? Can the inclusion of subjective performance pay helps us understand some observed phenomena in labor markets? To answer these important questions, this paper provides a theory of contract selection in a market setting.

In a seminal paper, MacLeod and Malcomson (1998, MM hereafter) provide the first model of contract selection between efficiency wages and subjective performance pay. The driving force in their model is market condition. In a market with more workers than jobs, a firm can always immediately and costlessly fill its vacancy after reneging on the promised bonus. Therefore, any subjective performance pay is not credible, and firms have to use efficiency wages solely to motivate workers. On the other hand, in a market with more jobs than workers, efficiency wages are useless in providing incentives because a worker can find another job immediately after being fired. As a result, firms use subjective performance pay solely to motivate workers.
However, market condition as a determinant of contract selection seems at odds with the fact there is unemployment in macroeconomy. Even if in some labor markets workers are temporarily on the short side, unemployed workers in other markets will enter into these markets. Because of this movement of labor and the fact that unemployment exists in macroeconomy, in the long-run workers are on the long side in any labor market. Therefore, market condition as a determinant of contract selection seems practically questionable. Moreover, in MM efficiency wages and subjective performance pay cannot be used together to motivate workers. In their performance pay equilibrium, bonuses of small magnitudes are never observed.\textsuperscript{1} This feature of discontinuity seems unpleasant. Finally, both efficiency wage equilibrium and performance pay equilibrium may exist at the same time in MM, but MM did not provide a rationale to select the more plausible equilibrium.

Complementary to MM, this paper provides a model of contract selection driven by exogenous turnover costs in labor markets. First, consistent with unemployment in macroeconomy, we assume that there is unemployment in each labor market thus ruling out market condition as a determinant of contract selection. Second, in our model firms are able to use combinations of efficiency wages and subjective performance pay to motivate workers, and the equilibrium bonus varies continuously with turnover costs. Third, given turnover costs there is a unique market equilibrium. Moreover, in our framework we are able to conduct welfare analysis, which is absent in MM. Finally, our model generates different empirical implications from MM.

Our basic model studies how turnover costs borne by firms affect contract selection. From firms’ point of view, efficiency wages are costly relative to subjective performance pay, because they have to pay a wage premium. However, subjective performance pay may not be credible due to the moral hazard problem on the part of firms: in labor markets with unemployment, a firm can immediately hire a new worker after reneging on the implicit bonus “promised” to its current worker. If a firm entails no turnover costs in hiring a new worker, then any subjective performance pay is not credible. On the other hand, if a firm entails some turnover costs in hiring a new worker, then some bonuses become credible: the firm can credibly enforce any bonus less than or equals to the turnover costs incurred in hiring a new worker. This is because now a worker can punish a reneging firm by

\textsuperscript{1}Specifically, the equilibrium bonus must be greater than the disutility of effort.
quitting. Since subjective performance pay is “cheaper” than efficiency wages in motivating workers, the optimal wage contract uses the maximum amount of bonus subject to the credibility constraint. Moreover, the higher the turnover costs, the bigger the optimal bonus.

The use of subjective performance pay not only reduces the required wage premium to motivate workers, but also reduces the total wage payment since it is “cheaper” than efficiency wages. Therefore, as turnover costs increase, the optimal bonus increases, the wage premium decreases, and the total wage payment decreases as well. These results hold as long as turnover costs are below some threshold (when the wage premium is positive). When turnover costs are in this range, firms use a combination of subjective performance pay and efficiency wages to motivate workers. As turnover costs increase further, no wage premium is required, thus firms use subjective performance pay solely to motivate workers.

After deriving the optimal contracts, we turn to study market equilibrium, which is determined by firms’ free entry condition. In market equilibrium, the revenue product of labor equals to the average labor cost, which consists of wage payment each period and the average turnover costs incurred per period on the equilibrium path. Interestingly, up to some threshold (when the wage premium is positive), an increase in turnover costs shifts the average labor cost curve downwards. This is because the reduction of wage payment per period is always bigger than the increase in average turnover costs incurred per period. Intuitively, an increase in turnover costs reduces the wage payment in each period, while the increase in average turnover costs per period is small since each job vacancy only incurs turnover costs once on the equilibrium path. This result implies that the equilibrium employment level is also increasing in turnover costs. Moreover, when the revenue product of labor is elastic enough, an increase in turnover costs leads to higher social welfare. This is a surprising result: a little bit friction in markets can be beneficial. The main reason behind this result is that friction in markets alleviates firms’ moral hazard problem and gives them commitment power, which in turn grants firms more flexibility in alleviating workers’ moral hazard problem.

Turnover costs in labor markets usually include three components: recruitment, exit and training costs. But not all of them have impacts on the optimal wage contract. According to MacLeod and Malcomson (1989), only the turnover costs that increase the rents of continued employment affect
relational contracts. Recruitment costs certainly fall into this category. Exit costs do so only when they are not merely transfers between firms and workers, so severance payments do not affect the optimal contract. Training costs create rents only when they are relationship specific.\footnote{There is a large literature on the issue of firm-specific human capital (for example, Hashimoto and Yu,1980).} Therefore, the turnover costs that are relevant to our model are mainly recruitment costs plus training costs that are related to firm specific skills.

In an extended model we incorporate workers’ search costs. Now wage contracts should not only motivate workers to exert effort (IC condition), but also induce unemployed workers to search (IR condition). Since there is unemployment in labor markets, an unemployed worker who actively searched, thus incurring search costs, may not find a job. It follows that inducing workers to search becomes more difficult as unemployment rate increases. On the other hand, motivating workers become more difficult as unemployment rate decreases. As a result, there is some cutoff value of employment level such that for employment levels higher than the cutoff the IC condition is binding, and for employment levels lower than the cutoff the IR condition is binding. If unemployed workers are able to perfectly adjust their search behavior according to market condition (are able to play mixed strategies regarding search), then the total wage payment is independent of employment level when the IR condition is binding. Thus, unlike the basic model where wages are always increasing in employment level when wage premium is positive, in the extended model wages are increasing in employment level only when employment is high, and is completely rigid when employment is low. This implies that wage-unemployment relationship changes over the course of business cycle: wages are procyclical in booms and rigid during recessions.

Like an increase in the turnover costs borne by firms, an increase in search costs decreases the necessary wage premium to motivate workers, though it has no impact on the optimal bonus. Moreover, an increase in either firms’ turnover costs or workers’ search costs expands the wage-rigidity region, and makes wages less sensitive to employment in the procyclical region.

Our model generates rich empirical implications. First, different labor markets (occupations) will adopt different forms of wage contracts. In particular, the component of efficiency wages or wage premium is negatively related to, and the amount of bonuses is positively related to, the
turnover costs borne by firms in labor markets. Second, workers paid by bonuses on average earn less than workers paid by fixed wages (efficiency wages). Third, occupations paid bonuses should have lower unemployment rates than occupations in which bonuses are seldom used. Fourth, wages are procyclical during booms, and are either rigid or countercyclical in recessions. Finally, the wage-unemployment elasticity is decreasing in turnover costs. All these predictions are consistent with some empirical evidence.

This paper is built on the literature of relational contracts (Bull, 1987; MM, 1989, 1998; Baker, Gibbons and Murphy, 1994; Levin, 2003). However, except for MM (1989; 1998) all the other papers study a one-firm-one-worker setting, hence both the firm’s and the worker’s outside options are exogenously given. Moreover, these papers do not study the contract selection between efficiency wages and subjective performance pay. MM (1989) offers a complete characterization of the set of relational contracts that can be implemented in a market setting, but it does not study the problem of contract selection.

Later shirking models of efficiency wages (Akerlof and Katz, 1989) incorporate a performance bond. But they implicitly assume that firms are able to commit: a firm never forfeit a worker’s bond if he does not shirk. In contrast to their assumption, our model, following the literature of relational contracts, assume that firms are not able to commit. The labor turnover models of efficiency wages (Salop, 1979; Stiglitz, 1974) treat the turnover rate as endogenous. They derive the result that firms with higher turnover costs may pay higher wages in order to reduce the turnover rate. This result is in direct contrast to the prediction of our model.

The structure of the paper is as follows. Section 2 sets up the basic model. Section 3 studies the optimal wage contracts. In Section 4, first the stationary market equilibrium is derived and then comparative statics and welfare analysis are conducted. In Section 5 we extend the basic model to incorporate search costs. Section 6 presents some empirical evidence that are consistent with the predictions of our model. Section 7 concludes the paper.
2 The Basic Model

Consider a market (an occupation) for basically homogenous labor with discrete periods. There are $L$ workers and many firms, which create $J$ job vacancies in total.\(^3\) While $L$ is exogenously given, there is free entry on the firms’ side; therefore $J$ is endogenous. Both workers and firms are risk neutral and share the same discount factor $\delta$. Each job has the same revenue product of labor $p$. Each firm takes $p$ as given, but in aggregate $p$ is a decreasing function of the total employment of non-shirking workers $E$, because the demand curve for products is downward-sloping. At the beginning of each period, unemployed workers and unfilled vacancies are randomly matched. Note that agents on the long side of the market may not get a match. At the end of each period, each existing job becomes nonprofiTABLE forever with probability $1 - \rho$ for exogenous reasons. This would result in exogenous separation if the job is filled in the current period. In any existing match that survives this shock, the worker and the firm simultaneously decide whether to continue the relationship next period. If either party decides to leave, then the match is broken up endogenously. All the agents without a match enter into the unmatched pool at the beginning of the next period.

If employed, a worker gets utility $W_t - eve_t$, where $W_t$ is the total wage payment and $e_t$ is his effort in period $t$. For simplicity, we assume that workers can either work ($e_t = 1$) or shirk ($e_t = 0$), with $v > 0$ as the disutility of work. The profit of a filled job vacancy in period $t$ is $pe_t - W_t$. Workers without job receive utility $u > 0$ per period, which can be interpreted as unemployment benefit. Consistent with employment at will, a firm’s commitment to the wage is legally binding for the current period of employment only, that is, only spot contracts are legally enforceable. Following incomplete contract literature, we assume that $e_t$ is observable but not verifiable.\(^4\) Therefore, spot contracts that are contingent on $e_t$ cannot be enforced by the court.

Nevertheless, since employment relationships have the potential to be long-term, firms may use implicit bonus. Specifically, the wage may include a fixed wage component and an end-of-period

\(^3\)As long as each firm has a small number of vacancies, the results of the model always hold.

\(^4\)An alternative specification would be that $e_t$ noisely determines the worker’s performance. While $e_t$ is not observable, the performance is observable but not verifiable. This implies that shirking cannot be detected for sure. The qualitative results of the paper still hold under this alternative specification.
bonus. In particular, the wage in period $t$ consists of a fixed wage $w_t \geq 0$ that the firm pays regardless of $e_t$ and a bonus $b_t \geq 0$ that the firm agrees to pay only if $e_t = 1$. While $w_t$ is legally enforceable, $b_t$ cannot be enforced by the court, hence it has to be self-enforcing.

Note that if the worker shirks, then the firm’s profit is non-positive. We assume that there is some employment level $\bar{E} > 0$ such that $p(\bar{E}) \geq u + v$; that is, for employment level $E \leq \bar{E}$, a firm earns a positive profit if the worker exerts effort. Therefore, firms will and have to motivate workers to exert effort. In addition, we assume that $p(L) \leq u + v$; that is, at full employment level, the revenue product of labor is less than the opportunity cost of labor. This implies that $J \leq L$, i.e., workers are always on the long side of the market.

There are turnover costs in the labor market. Firms incur recruiting costs in finding job candidates. Workers incur search costs in finding job openings. Moreover, each firm may require a skill which is firm-specific. For simplicity, we assume that it takes one period for a new employee to acquire the firm-specific skill. During that period the output of a new worker is less than that of an old worker by $c_F \geq 0$. The parameter $c_F$ captures the degree of firm-specificity of jobs: the more firm-specific jobs are in an occupation, the larger the $c_F$. We denote $c = c_F + c_R$ ($c_R$ represents recruiting costs) as the turnover costs. For simplicity, we assume that $c_F$ is borne by the firm alone. Later on we will discuss how relaxing this assumption affects the results of the model. And in the basic model we ignore workers’ search costs, which will be studied later in section 5. For simplicity, we assume that if a worker is separated from a firm, his skill which is specific to that firm degenerates immediately. Therefore, all workers in the unmatched pool are homogenous regardless of their employment history.

This is essentially an infinitely repeated game, with some employment relationships reshuffled in

---

5By imposing $w_t \geq 0$ and $b_t \geq 0$, we implicitly assume that workers are subject to limited liability.

6For some jobs, firms are willing to pay employment agencies large sums (on the order of one to four months’ salary) for finding workers (Lang, 1991).

7Another interpretation is that it takes one period for a firm to train a new worker to get the firm-specific skill, and the training cost is $c_F$.

8This assumption is not unrealistic. An empirical study by Green et al (2000) shows that expenses on most transferable training are paid by employers.
each period. The timing of a typical stage game is shown in figure 1. At the beginning of period $t$, new vacancies are created. Then unemployed workers and unfilled job vacancies are randomly matched. In each match, the firm offers the worker a contract. The worker can either accept or reject the offer. If rejects the offer, the worker is unemployed for the whole period. Then each employed worker chooses effort. Afterwards employed workers get paid. After payment has been made, a $1 - \rho$ fraction of current jobs die and workers matched with those jobs are exogenously separated. In any existing match that survives this shock, the worker and the firm simultaneously decide whether to continue the relationship in next period. If either party decides to leave the current relationship, the match breaks up. All the unmatched agents enter into the unmatched pool in the beginning of next period.

Matching of jobs with workers in the unmatched pool

<table>
<thead>
<tr>
<th>Firm pays bonus $b_t$ or renege</th>
</tr>
</thead>
<tbody>
<tr>
<td>Separation decisions</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Period $t+1$</th>
</tr>
</thead>
<tbody>
<tr>
<td>Workers make effort $e_t$</td>
</tr>
<tr>
<td>Each job becomes unprofitable with Prob $1-\rho$</td>
</tr>
</tbody>
</table>

Figure 1: Timing

Note that there are double moral hazard problems: workers may want to shirk and collect the fixed wage $w_t$; firms may want to renege the bonus $b_t$ after workers exerting effort. The concern for external reputation may mitigate the moral hazard if market can observe accurately which party is at fault in case of separation. However, we rule out this effect due to the following reasons. First, outsiders don’t know exactly whether a separation occurs due to exogenous reasons or cheating. Second, even if a separation is known due to cheating, outsiders don’t know exactly which party is at fault. In other words, we assume that the word of mouth in labor markets are not reliable.

The model is similar to MM, but with two major departures. First, we introduce turnover costs in labor markets, which is not studied in their model. It turns out that optimal contracts crucially depend on turnover costs. Second, we only consider the case $J \leq L$, while in their model both cases
$J \leq L$ and $J > L$ are studied. Actually, market condition is the driving force of contract selection in their model.\textsuperscript{9}

There are many relational contracts that can be supported as perfect equilibria, some of them may involve complicated strategies.\textsuperscript{10} To support equilibrium, the party who violates the relational contract should be punished by receiving a lower continuation payoff. But there are upper bounds for the punishment. This is because both parties have outside options: after separating from the current partner each party can find another partner in the future. For most of the paper, we restrict our attention to trigger strategies. In particular, if the worker shirks, the firm will fire the worker at the end of the period. Similarly, if the firm reneges on bonus (pay less than $b_t$), the worker will not accept any bonus offered by the firm in the future: he either quits or the firm retains him by offering a fixed wage contract which gives him more utility than his outside option. Later on we will show that after reneging the firm always let the worker quit. So without loss of generality the worker will always quit if the firm reneges. These strategies are appealing in practice: if at least one party violates the relational contract, the informal relationship becomes sour and it is unlikely to continue.

### 3 The Optimal Contracts

**Value functions** We are interested in stationary equilibrium, in which employment level, the fixed wage and implicit bonus are constant over time. So we can drop all the time subscripts. Denote $U^N$ ($U^S$) as the expected discounted lifetime utility of an employed non-shirker (shirker), and $\bar{U}$ as the expected discounted lifetime utility of an unemployed worker. Let $E$ be the number of employed workers in a market, and $a$ be the job acquisition rate. Note that both $E$ and $a$ are endogenous.

\textsuperscript{9}Another minor difference is that in MM there is a job creation cost. The impact of the job creation cost is fundamentally different from those of turnover costs. Specifically, the presence of job creation cost would not affect firms’ incentive to renege, thus having no impact on the optimal contracts. On the other hand, the presence of turnover costs would reduce firms’ incentive to renege, thus affecting the optimal contracts (which will be shown later). The job creation cost only affects firms’ zero profit condition. In our model, we can also introduce a job creation cost. But it merely pushes up firms’ average cost curve by a constant; the qualitative results of our model remain unaffected.

\textsuperscript{10}For a complete characterization of the set of perfect equilibria induced by relational contracts, see MM (1989).
variables. The value function $U^N$ is the following (suppose the firm does not renege):

$$U^N(w, b) = (w + b - v) + \delta \rho U^N(w, b) + (1 - \rho)U(w, b)$$  \hspace{1cm} (1)

The value of $U^N$ consists of two terms: the current period payoff and continuation value. The current period payoff for a non-shirker is his wage payment $w + b$ minus his disutility of effort $v$. With probability $1 - \rho$ the current job will die in the next period so the worker will enter into the unemployment pool at the beginning of the next period. In this case his continuation value is $\delta U(w, b)$. With probability $\rho$ the current job will survive in the next period and the worker’s continuation value is $\delta U^N(w, b)$. Similarly, other value functions are:

$$U^S(w, b) = w + \delta U(w, b)$$ \hspace{1cm} (2)

$$\overline{U}(w, b) = a \cdot \max\{U^N(w, b), U^S(w, b)\} + (1 - a)[u + \delta U(w, b)]$$ \hspace{1cm} (3)

If an employed worker shirks, the current relationship breaks up for sure at the end of the period, so the continuation value is $\delta \overline{U}(w, b)$. For an unemployed worker (at the beginning of a period), with probability $a$ he will find a job in that period and his continuation value is $\max\{U^N(w, b), U^S(w, b)\}$; with probability $1 - a$ he will stay unemployed in that period and his continuation value is $u + \delta U(w, b)$.

Similarly, denote $V^N$ ($V^R$) as the expected discounted lifetime profit of a filled job vacancy if the firm does not renege (reneges), and $\overline{V}$ as the expected discounted lifetime profit of an unfilled job vacancy. The value functions are the following (suppose the worker does not shirk):

$$V^N(w, b) = (p - w - b) + \delta \rho V^N(w, b)$$ \hspace{1cm} (4)

$$V^R(w, b) = (p - w) + \delta \rho \overline{V}(w, b)$$ \hspace{1cm} (5)

$$\overline{V}(w, b) = \max\{V^N - c, V^R - c\}$$ \hspace{1cm} (6)

\footnote{Given workers are playing trigger strategy, for a reneging firm paying 0 instead of any positive bonus less than the promised $b$, is the most profitable deviation.}
The continuation value of $V^N$ has two components. With probability $1 - \rho$, the job dies in the next period, and the firm gets 0 continuation value; with probability $\rho$, the job survives in the next period, and the firm’s continuation value is $\delta V^N(w, b)$. $V^R$ is different from $V^N$ in continuation value: after the firm reneges, the worker quits and the continuation value is $\delta \rho \overline{V}(w, b)$ instead of $\delta \rho V^N(w, b)$. $\overline{V}$ is the firm’s fallback position if it reneges. Since there is unemployment in the labor market, the firm can always immediately find a new worker to fill its vacancy, by incurring turnover cost $c$. Therefore, $\overline{V}(w, b) = \max\{V^N, V^R\} - c$.

**Programming problem** Each firm offers a relational contract to maximize its profit, taking the employment $E$ (hence $a$) as given. The relational contract should satisfy the following conditions. Unemployed workers should be willing to accept the contract, and employed workers should have incentives to exert effort. Firms should have incentive to create vacancies, and firms should not have incentives to renege. Note that maximizing profit is equivalent to minimizing total wage payment $W$. Mathematically, the programming problem is as follows

\[
\begin{align*}
\min_{w \geq 0, b \geq 0} & \quad W = w + b \\
\text{subject to} & \\
U^N & \geq \overline{U} \quad \text{(IRW)} \\
\overline{V} & \geq 0 \quad \text{(IRF)} \\
U^N & \geq U^S \quad \text{(ICW)} \\
V^N & \geq V^R \quad \text{(ICF)}
\end{align*}
\]

The IRF condition depends on the revenue product of labor $p(E)$. At time being, we ignore this condition and discuss it in the next section when we study market equilibrium. Substitute equation (1) and (2), the ICW (No-shirking) condition becomes

\[
\delta \rho (U^N - \overline{U}) \geq (v - b) \quad (7)
\]

The left hand side of (7) is the rent of continued employment shared by the worker, and the right
hand side is the current period gain from shirking. Note that efficiency wages (measured by $U^N - \overline{U}$) and performance pay $b$ are substitutes in motivating workers: either the bonus $b$ should be big enough to reduce the gain of shirking, or the worker’s rent from continued employment, which is created by paying efficiency wages, should be big enough. Note that the firm can also use a combination of both.

Similarly, the ICF (No-reneging) condition can be simplified as:

$$b \leq \delta \rho c$$

(8)

The left hand side of (8) is the firm’s current period gain from reneging; the right hand side is the expected cost of reneging in the next period. Since it can always have a job vacancy filled immediately, the firm’s rent from continued employment comes solely from the fact that retaining an old worker saves turnover costs $c$. Therefore, any credible $b$ has an upper bound $\delta \rho c$. Any bonus greater than this upper bound is non-credible since the firm has an incentive to renge.

**Pure efficiency wages** First we consider the case $c = 0$. From the No-reneging condition (8), $b$ has to be 0 since any positive bonus is not credible. As a result, performance pay is not available at all. This corresponds to the situation studied by Shapiro and Stiglitz (1984). Firms have to use solely efficiency wages to motivate workers: a positive rent of $U^N - \overline{U}$ has to be created through paying efficiency wages and the threat of firing serves as a discipline device.

Given $b = 0$, the programming problem is simplified as

$$\min_{w \geq 0} w$$

subject to

$$U^N \geq \overline{U} \quad \text{(IRW)}$$

$$\delta \rho (U^N - \overline{U}) \geq v \quad \text{(ICW)}$$

Note that the IRW condition is redundant. The ICW condition can be further simplified as:

$$w \geq u + \frac{v}{(1-a)\delta \rho}$$

(9)

12 In MM, a firm’s rent from continued employment comes from the fact that the firm may not have its vacancy filled immediately, which is possible only when $J > L$. This rules out the use of performance pay when $J \leq L$. 

13
In the optimal contract, condition (9) is obviously binding. Therefore, the optimal wage contract has the following form:

\[ b^* = 0; \ w^* = u + \frac{v}{(1 - a)\delta \rho} \]

Note that the IRW condition is slack in the optimal contract. Define the wage premium \( w^p \) as the extra utility per period enjoyed by an employed worker, i.e., \( w^p \equiv (b^* + w^* - v) - u \). Under the optimal contract, an employed worker is enjoying a positive wage premium:

\[ w^p = \frac{1 - (1 - a)\delta \rho}{(1 - a)\delta \rho} \]

**Combinations of efficiency wages and performance pay**  With a positive \( c \), firms have more freedom to choose \( b \) in relational contracts. From No-reneging condition (8), the credible \( b \) the firm can choose is in the range \([0, \delta \rho c]\). The bigger the \( c \), the more freedom the firm has to choose \( b \). To facilitate analysis, we first prove a lemma.

**Lemma 1** If \( c \in [0, \frac{v}{\delta \rho}) \), in the optimal contract the firm should set \( b^* = \delta \rho c \). If \( c \geq \frac{v}{\delta \rho} \), in the optimal contracts the firm can set any \( b^* \in [v, \delta \rho c] \) subject to \( w^* = u + v - b^* \geq 0 \).

**Proof.** After some algebra, the programming problem can be simplified as the following:

\[
\min_{w \geq 0, b \geq 0} \{ w + b \}
\]
\[ w + b \geq u + v \quad \text{(IRW)} \] (10)
\[ w \geq u + k(v - b) \quad \text{(ICW)} \] (11)
\[ b \leq \delta \rho c \quad \text{(ICF)} \] (12)

where \( k = \frac{1}{(1 - a)\delta \rho} > 1 \). Condition (11) can be further reformulated as

\[ w + b \geq u + v + (k - 1)(v - b) \] (13)

From (13), we see that if \( b < v \), only (11) is binding, and the total wage payment \( w + b \) is decreasing in \( b \); so \( b \) should be set as high as possible subject to condition (12). Condition (10) is binding if and only if \( b \geq v \). Actually, condition (10) (IRW) specifies the low bound of total wage payment, which
is $u + v$. Therefore, when $c < \frac{u}{\delta \rho}$, the optimal contract is: $b^* = \delta \rho c$, and $w^* = u + k(v - b^*)$. When $c \geq \frac{u}{\delta \rho}$, the optimal contracts are: $b^* \in [v, \delta \rho c]$ subject to $w^* = u + v - b^* > 0$. ■

The intuition for lemma 1 is the following. There are two ways to motivate workers: efficiency wages and performance pay. Performance pay is less costly than efficiency wages from firms’ perspective. This is because performance pay discourages shirking directly with no need to increase total wage payment, while efficiency wages entail firms to increase total wage payment. However, performance pay is restricted by the moral hazard problem on the part of the firm: it may renege on the bonus if it is too high. This moral hazard can be alleviated by the presence of turnover costs $c$: the firm will incur $c$ in hiring a new worker next period if it reneges on bonus. The upper bound of credible bonuses thus is increasing in turnover costs $c$. When $\delta \rho c < v$, the firm should set the highest credible bonus $b^* = \delta \rho c$ to reduce the necessary wage premium required to motivate workers. When $\delta \rho c \geq v$, setting any $b \in [v, \delta \rho c]$ is enough to motivate workers, and the firm does not need to pay wage premiums. But now the IRW condition is binding, so the total wage payment cannot be reduced further.

We can calculate the wage premium $w^p$ in the optimal contracts. From lemma 1,

$$w^p = (b^* + w^* - v) - u = (k - 1) \max\{v - \delta \rho c, 0\}$$

From (14) it is clear that the wage premium is decreasing in $c$. Moreover, when $c \geq \frac{u}{\delta \rho}$ the wage premium equals to 0.

The following proposition summarizes the previous discussions:

**Proposition 1** If $c \geq \frac{u}{\delta \rho}$, the optimal contracts have the following form: $b^* \in [v, \delta \rho c]$ subject to $w^* = u + v - b^* \geq 0$; workers receive no wage premium; the optimal contracts are purely in the form of performance pay. If $c \in \left[0, \frac{u}{\delta \rho}\right)$, the optimal contract is the following: $b^* = \delta \rho c$ and $w^* = u + \frac{v - \delta \rho c}{(1 - a) \delta \rho}$; employed workers receive a positive wage premium, which is decreasing in $c$; the optimal contract is a combination of performance pay and efficiency wages, and the efficiency wage component is decreasing in $c$ while the performance pay component is increasing in $c$. When $c = 0$, the optimal contract has no performance pay component and is in the form of pure efficiency wages.
**Robustness** Recall that we have assumed that the worker always quits after the firm reneges the bonus. But trigger strategy only specifies that the worker will not accept any bonus offered by the firm if it reneges. If the reneging firm offers a pure efficiency wage contract (without bonus), the worker may be willing to accept it, especially when employed workers enjoy a wage premium. Note that if employed workers enjoy no wage premium, quitting is a weakly dominant strategy. So we should only worry about the case in which wage premium is positive, i.e., $c \in [0, \frac{v}{\delta \rho})$. Denote $V^{EF}$ as the firm’s expected profit if it offers the optimal pure efficiency wage contract to retain the old worker. On the other hand, if the firm hires a new worker from the market next period, its expected profit is $V^N - c$.

$$V^N - c - V^{EF} = \frac{p - u - v - (k - 1)(v - \delta pc)}{1 - \delta \rho} - c - \frac{(p - u - kv)}{1 - \delta \rho}$$

Therefore, the firm always let the old worker quit after reneging.

So far we have assumed that $c$ is borne by the firm solely. Now we briefly discuss what happens if we relax this assumption. Suppose that the firm and the new worker share turnover costs $c$ according to some bargaining rule, with the firm bearing cost $\theta c$ and the new worker bearing $(1 - \theta)c$ ($\theta \in (0, 1)$). We can go through the same analysis again. Now the firm can enforce a credible bonus only up to $\theta(\delta pc)$, which is less than $\delta pc$ in the basic model. However, now the worker has less incentive to shirk since if he shirks he will be fired and bear the additional cost $(1 - \theta)c$ in case he finds a new job next period. The amount of necessary efficiency wages is still decreasing in $c$. Overall, the optimal bonus is decreased, but the amount of efficiency wages is more or less the same. The qualitative results of the model remain the same.

Note that any endogenously created turnover costs has no value in overcoming the moral hazard problem. Suppose that an employment contract specifies that the firm pays the worker $c$ whenever separation occurs. This severance pay $c$ enables the firm to credibly enforce a bonus $c$. However, under this circumstance the worker’s moral hazard problem is not altered: though the worker does not receive the bonus $c$ if he shirks, he receives a severance pay $c$ instead, thus there is no punishment for shirking. Similarly, any severance pay from the worker to the firm has no value in overcoming
the moral hazard problem either.\footnote{In terms of the theory of relational contracts, for a relational contract to be self enforcing the surplus generated inside the relationship has to be bigger than those generated outside the relationship. Without exogenous turnover costs, the surplus generated inside the relationship always equals to those generated outside the relationship; therefore, no implicit agreement can be self enforcing.}

With $c \in (0, \frac{1}{\delta\rho})$ (so employed workers enjoy a positive wage premium), the trigger strategies associated with the optimal contract do constitute a subgame perfect Nash equilibrium. This is because the worker and the firm make their separation decisions simultaneously. Under this assumption, after one party’s deviation, the strategy profile (quit, fire) is optimal since unilateral deviation would not change the outcome. Are these strategies renegotiation proof? They are definitely not renegotiation proof, since any endogenous separation results in welfare loss (turnover costs $c$ has to be incurred again). Fortunately, if the optimal contract can be supported by trigger strategies, we are able to construct another equilibrium which is strongly renegotiation-proof in the sense of Farrell and Maskin (1989) to support the optimal contract.\footnote{Farrell and Maskin (1989) define strongly renegotiation proof equilibrium as an equilibrium in which none of its continuation equilibria is strictly dominated by another equilibrium.} The basic idea is to keep the employment relationship going always, punish the party who violates the contract, and reward the other party.

Proposition 2 Let $(w^*, b^*)$ be the optimal contract derived under trigger strategies, then there exist strategies which can support the optimal contract and constitute a strongly renegotiation proof equilibrium.

Proof. See the Appendix. $\blacksquare$

The above strategies specify an optimal punishment scheme possibly at the expense of limited liability. In the cooperative phase, the worker gets the surplus $U^N - \bar{U}$, the firm gets the surplus $V^N - \bar{V}$. If one party deviates, the deviating party will get a continuation payoff equals to his outside option, and the other party is rewarded with the whole surplus. By keeping the relationship going and rewarding the punishing party, the incentive constraints remain the same as in the trigger strategy equilibrium, but the punishing party incurs no loss in carrying out punishment.
4 Market Equilibrium

Stationary equilibrium  In market equilibrium, all the firms in the market offer the same optimal contract derived in the previous section, since they all face the same programming problem. Moreover, in market equilibrium all the employed workers exert effort; all the firms pay the promised implicit bonus (if there is any); and employment relationships end only due to exogenous separation. We are interested in stationary market equilibrium, in which the number of newly created job vacancies equals to the number of jobs that die in every period. In other words, the equilibrium employment level is constant over time. Mathematically, the employment level in stationary equilibrium is defined by

\[ (1 - \rho)E = a(L - \rho E) \]

So the job acquisition rate \( a \) in stationary equilibrium is

\[ a = \frac{(1 - \rho)E}{L - \rho E} \quad (15) \]

Note that \( a \in [0, 1] \) because \( E \leq L \). Moreover, \( a \) is an increasing function of \( E \).

Since there is free entry on firms' side, at the equilibrium employment level \( E^* \) any new job vacancy’s profit

\[ \nabla(w^*, b^*, E^*) = \frac{p(E^*) - w^* - b^*}{1 - \delta \rho} - c = 0 \quad (16) \]

Alternatively, (16) can be written as

\[ p(E^*) = w^* + b^* + (1 - \delta \rho)c \quad (17) \]

The left hand side of (17) is the revenue product of labor per period, which specifies a revenue product curve of labor. The right hand side of (17) specifies the average labor cost (ALC) per period, which consists of two terms: wage payment per period \((w^* + b^*)\) and per period turnover costs spreading over the expected life time of a job \(((1 - \delta \rho)c\). Specifically, for \( c \in [0, \frac{v}{\delta \rho}] \)

\[ ALC(E) = w^* + b^* + [1 - \delta \rho]c = u + c + \frac{v - \delta \rho c}{(1 - a(E))\delta \rho} \]

\[ = u + \frac{v}{(1 - a(E))\delta \rho} - \frac{a(E)}{1 - a(E)}c \]

\[ ALC(E) = u + \frac{v}{(1 - a(E))\delta \rho} \quad (18) \]

For \( c \geq \frac{v}{\delta \rho} \)

\[ ALC(E) = w^* + b^* + (1 - \delta \rho)c = u + v + (1 - \delta \rho)c \]

\[ ALC(E) = u + v + (1 - \delta \rho)c \quad (19) \]
Note that the $ALC$ is a function of $E$. Therefore, the right hand side of (17) specifies a ALC curve. The stationary market equilibrium is determined by the intersection of the revenue product curve and the ALC curve.

By assumption, the revenue product curve $p(E)$ is downward sloping. When $c \in [0, \frac{v}{\delta p})$, from (18) we can see that $ALC$ is increasing in $E$ because $\frac{\partial ALC}{\partial a} > 0$ and $a$ is increasing in $E$. Therefore, the ALC curve is upward sloping. Note that when $E$ is close to the full employment level $L$, $a$ converges to 1, and $ALC$ goes to infinity. So full employment cannot be achieved if firms need to pay efficiency wages. On the other hand, when $c \geq \frac{v}{\delta p}$, from (19) we can see that $ALC$ is independent of $E$, thus the ALC curve is horizontal.

Figure 2: Market Equilibrium

Figure 2 illustrates the determination of market equilibrium for the case $c \in [0, \frac{v}{\delta p})$. Point $A$, the intersection of the $p(E)$ curve and the ALC curve, specifies the market equilibrium. At point $A$, firms are offering the optimal contract and each new job vacancy earns zero profit. Note that the market equilibrium is unique. For any other point in figure 2, firms will either earn a profit other than zero or they are not offering the optimal contract. Notice that point $A$ specifies equilibrium employment level $E^*$ and the optimal wage contract simultaneously.

**Comparative statics** In the following we are going to study how changes in $c$ affect the market equilibrium. Consider three labor markets, which are the same (have same parameter values) except
that they have different turnover costs $c$. In particular, $c_3 = \frac{\nu}{\delta \rho}$, $c_2 \in (0, \frac{\nu}{\delta \rho})$ and $c_1 = 0$. According to (18), $ALC$ is decreasing in $c$. Therefore, at any employment level, $ALC(c_1) > ALC(c_2) > ALC(c_3)$. Correspondingly, the relative positions of the ALC curves are the following: $ALC(c_1)$ lies above $ALC(c_2)$, which lies above $ALC(c_3)$. Note that $ALC(c_1)$ and $ALC(c_2)$ are upward sloping while $ALC(c_3)$ is horizontal. Figure 3 illustrates the market equilibria for different markets. In the figure, points $A$, $B$ and $C$ represent the market equilibrium for market 1, 2 and 3 respectively. The equilibrium employment levels are increasing in $c$: $E^*_1 < E^*_2 < E^*_3$.

![Figure 3: Comparative Statics](image)

In general, for $c \in [0, \frac{\nu}{\delta \rho}]$ the equilibrium employment level $E^*$ is increasing in $c$, because an increase in $c$ shifts the ALC curve downwards. The intuition for this result is the following. An increase in turnover costs by $\Delta c$ can decrease the total wage payment in every period by $(\frac{1}{1-a} - 1)\delta \rho \Delta c$. On the other hand, on the equilibrium path each job vacancy incurs turnover costs only once (in the first period). Spreading over a new job vacancy’s expected life time, the average turnover costs per period increases by $(1-\delta \rho)\Delta c$. Overall, the first effect dominates the second one. Changes in $c$ also affect the slope of the ALC curve. By (18) an increase in $c$ makes $ALC$ increases more slowly as $E$ increases. Put it another way, an increase in $c$ makes the ALC curve flatter.

But this result is reversed when $c \geq \frac{\nu}{\delta \rho}$. If $c$ falls in this region, by (19) an increase in $c$ shifts the ALC curve upwards. This is because an increase in $c$ cannot reduce the wage payment further.
(the IRW condition is binding), but it directly pushes up the average turnover costs. Hence, the equilibrium employment level \( E^* \) is decreasing in \( c \). The following proposition summarizes the above analysis.

**Proposition 3** For \( c \in [0, \frac{v}{\delta \rho}] \), an increase in \( c \) shifts the ALC curve downwards and makes the ALC curve flatter; the equilibrium employment level \( E^* \) is increasing in \( c \). For \( c \geq \frac{v}{\delta \rho} \), an increase in \( c \) shifts the ALC curve upwards; the equilibrium level \( E^* \) is decreasing in \( c \).

By (18) and (19), \( ALC \) is increasing in the exogenous turnover rate \( (1 - \rho) \). The intuition for this result is the following (for \( c \in [0, \frac{v}{\delta \rho}] \)). First, the higher the exogenous turnover rate, the lower the expected saving in turnover costs if a firm does not renege, which means that the upper bound of the credible bonuses is smaller. Second, the higher the exogenous turnover rate, the lower the rent of continued employment shared by a non-shirker. This is because he now has a higher probability entering into the unemployment pool next period even if he does not shirk. Third, the job acquisition rate in stationary equilibrium becomes higher as \( 1 - \rho \) increases, since the fact that more jobs die means more new vacancies are created in stationary equilibrium; hence the wage premium required to motivate workers is higher. Finally, the average turnover costs per period are increasing in \( 1 - \rho \), since a new job’s expected life time is increasing in \( \rho \). All these effects work in the same direction and result in a high \( ALC \). Therefore, other things equal, an increase in the exogenous turnover rate shifts the ALC curve upwards. Consequently, it decreases the equilibrium level of employment.

Similarly, by (18) and (19) \( ALC \) is decreasing in discount factor \( \delta \). This result is intuitive: as future becomes more important, repeated interaction becomes more effective in disciplining opportunistic behavior, thus average labor cost decreases. More specifically, an increase in discount factor shifts the ALC curve downwards, thus increases the equilibrium employment level.

**Welfare properties** Now we study how changes in turnover costs \( c \) affect social welfare. In the market equilibrium, each firm’s expected profit is 0, since \( p(E^*) = ALC(E^*) \). The social value of the total output is \( \int_0^{E^*} p(E)dE \), which is consumers’ total willingness to pay for the total output.

\(^{15}\)When \( c > \frac{v}{\delta \rho} \), only the last effect remains. But still a higher exogenous turnover rate results in a higher \( ALC \).
The total social surplus $S$ (per period) of a market equilibrium with employment $E^*$ is

$$S(E^*, c) = \int_0^{E^*} p(E) dE - E^*(u + v) - (1 - \rho)E^*c$$

(20)

The second term is the total cost of labor, and the third term is total turnover costs incurred per period. Take derivative of (20) with respect to $c$,

$$\frac{dS}{dc} = [p(E^*) - (u + v) - (1 - \rho)c] \frac{\partial E^*}{\partial c} - (1 - \rho)E^*$$

(21)

We first consider the case $c \in [0, \frac{v}{\delta\rho})$. Using equations (17) and (18), we can get $\frac{\partial E^*}{\partial c}$ and substitute it into (21)

$$\frac{dS}{dc} = \frac{v}{\delta\rho} - (1 - a)\delta v - (1 - \rho)\delta pc - ap\delta pc \frac{a}{-p'(E^*)(1 - a)^2 + (\frac{v}{\delta\rho} - c)a'} - (1 - \rho)E^*$$

Where prime denotes derivative with respect to $c$. Using the stationarity condition (15),

$$\frac{dS}{dc} = (1 - \rho)E^*[\frac{v}{\delta\rho}(L - \rho E^*) - (L - E^*)v - (1 - \rho)Lc}{-p'(E^*)(L - E^*)^2 + (\frac{v}{\delta\rho} - c)(1 - \rho)L} - 1$$

After some simplification,

$$\frac{dS}{dc} \geq 0 \iff 1 - \delta \frac{v}{\delta\rho} \geq -p'(E^*)(L - E^*)$$

(22)

Condition (22) is satisfied if $|p'(E^*)|$ is small enough. In other words, if the revenue product of labor is elastic enough, then the social surplus is increasing in turnover costs $c$. Intuitively, an increase in $c$ has two opposite effects on social welfare (see equation (21)). On one hand, it directly increases the total turnover costs, thus reducing social welfare. On the other hand, it increases the equilibrium level of employment, thus leading to higher social welfare. If the revenue product of labor is elastic enough, a small increase in $c$ can induce a big increase in equilibrium employment level, thus the positive effect dominates the negative effect, and social welfare increases. If the revenue product of labor is linear, then from (22) we can see that an increase in $c$ is more likely to increase social welfare if the initial equilibrium employment level $E^*$ is higher.

When $c \geq \frac{v}{\delta\rho}$, the social surplus is decreasing in $c$. This is because an increase in $c$ reduces the equilibrium employment $E^*$, thus both terms are negative in equation (21). So we proved the following proposition.
Proposition 4  If the revenue product of labor is elastic enough, the social surplus is increasing in $c$ for $c \in [0, \frac{\delta \rho}{\sigma \rho})$. When $c \geq \frac{\delta \rho}{\sigma \rho}$, the social surplus is decreasing in $c$.

Proposition 4 is a surprising result: a little bit friction in markets can improve social welfare. Common wisdom tells us that friction in markets is always bad, since it impedes the smooth functioning of markets. However, our model shows that if there is a double moral hazard problem in the market, a little bit friction in the market actually can make the market function more effectively. The main reason is that without exogenous friction contingent contracts are not available, thus to motivate one side of the market (workers) certain amount of friction has to be created endogenously (by using efficiency wages). The presence of some exogenous friction alleviates firms’ moral hazard problem and gives them commitment power, which makes contingent contracts (performance pay) feasible. Contingent contracts not only reduce the necessary amount of endogenously created friction to motivate workers, but also reduce the total amount of friction in the market that are necessary to motivate workers. It is the flexibility of contingent contracts, which are made possible by the presence of exogenous friction, that reduces the total amount of inefficiency.

**Empirical predictions**  Our model generates several testable empirical implications. The first implication is about forms of employment contracts. Our model predicts that labor markets with different turnover costs will use different forms of employment contracts. In particular, occupations with high turnover costs are paid a high bonus, and those with low turnover costs are paid a low bonus. The second implication is about total wage payment. The model predicts that occupations with high turnover costs are paid a low total wage, and those with low turnover costs are paid a high total wage. This is because high turnover costs lead to high bonuses, which reduce the wage premium. The second result also implies that workers paid higher bonuses actually earn less than those paid low bonuses.

The third implication is about the relationships among turnover costs, bonuses, and unemployment. The model predicts that occupations paid higher bonuses should have lower levels of unemployment. However, the relationship between turnover costs and unemployment is non-monotonic. Among the occupations with low turnover costs, the occupations with higher turnover costs should
have lower levels of unemployment. On the other hand, among the occupations with high turnover costs, the occupations with higher turnover costs should have higher levels of unemployment. The final implication is about the sensitivity of wage payment to employment levels. The higher the turnover costs, the lower the wage premium, thus wage increases more slowly as employment increases. This implies that wages are less sensitive to employment level in occupations with high turnover costs than in occupations with low turnover costs.

These predictions are different from those in MM. First, in their model either pure efficiency wages or pure performance pay is used to motivate workers (firms cannot use a combination of both). In particular, occupations with unemployment cannot be paid any bonus. In our model, firms can use a combination of both to motivate workers: the amount of bonuses varies continuously with turnover costs $c$. Second, their model predicts that contract selection is determined by market condition, while our model predicts that the form of employment contracts is determined by turnover costs.

5 Relational Contracts with Search Costs

In the basic model, we have assumed that workers incur no cost in searching for jobs. More realistically, workers do incur search costs. Search costs may include the effort to find open vacancies, to prepare applications, and to prepare interviews, etc. In this section we extend the basic model to incorporate workers’ search costs. Search costs are directly borne by workers. More importantly, a worker incurs search costs as long as he has actively searched for a job, regardless of whether he finds a job. As we will see later, this feature has qualitative impact on the properties of market equilibrium. Although workers bear search costs directly, firms have to induce workers to search for two reasons. First, firms need workers to fill new vacancies. Second, to more effectively discipline shirkers firms have incentives to reduce the effective job acquisition rate by inducing workers to search.

We model unemployed workers’ search behavior as follows. Anticipating the wage contracts that firms are going to offer, in each period each unemployed worker decides whether and with which probability to search (here we allow workers to play mixed strategy regarding search). If a worker
searches, he incurs search costs $c_S > 0$ in that period regardless of the outcome. After unemployed workers made their search decisions, the effective job acquisition rate $a$ is determined, which is the ratio of the number of unfilled vacancies to the population of unemployed workers who search actively. Since the environment is symmetric for all unemployed workers, we focus on symmetric strategies: each unemployed worker searches with the same probability $\sigma \in [0, 1]$. Given $\sigma$, the stationary job acquisition rate

$$a = \min\{1, \frac{(1 - \rho)E}{\sigma(L - \rho E)}\}$$

The presence of search costs $c_S$ has two effects on wage contracts. First, it can discipline employed workers: if they shirk, they are going to be fired and have to incur search costs $c_S$ for several periods to find another job. Second, now firms have to induce workers to search, since search costs are directly borne by workers. Specifically, with the presence of $c_S$, the value functions become the following.

$$U^N = (w + b - v) + \delta[\rho U^N + (1 - \rho)\bar{U}]$$

$$U^S = w + \delta \bar{U}$$

$$\bar{U} = \max\{u + \delta \bar{U}, a \ast \max\{U^N, U^S\} + (1 - a)(u + \delta \bar{U}) - c_S\}$$

Note that $\bar{U}$ is the maximum of two payoffs: the payoff if a worker searches and the payoff if he does not search. Firms’ value functions are still the same as those in the basic model.

Unlike the basic model, here the effective job acquisition rate $a$ depends on unemployed workers’ search behavior. It is easy to check that $a$ reaches its low bound when $\sigma = 1$. Define this low bound as

$$a = \frac{(1 - \rho)E}{L - \rho E}$$

16 This restriction is not essential to the results of the model. In any mixed strategy equilibrium (whether symmetric or not), each unemployed worker should be indifferent between searching and not searching. This is the only feature that is essential to our model.
Specifically, \( a \) is determined by the following formula:

\[
a = \begin{cases} 
  a & \text{if } a[\max\{U^N, U^S\} - (u + \delta U)] - c_S \geq 0 \quad (C1) \\
  0 & \text{if } \max\{U^N, U^S\} - (u + \delta U) - c_S < 0 \quad (C2) \\
  \frac{c_S}{U^N - (u + \delta U)} & \text{if neither (C1) nor (C2) is satisfied}
\end{cases}
\]

In the first case, each unemployed worker searches with \( \sigma = 1 \). This is because the return of search is higher than search costs even if all the unemployed workers search with \( \sigma = 1 \). In the second case, each unemployed worker does not search \( (\sigma = 0) \). This is due to the fact that the return of search is lower than search costs even if the job acquisition rate \( a = 1 \). In the third case, neither condition (C1) nor condition (C2) is satisfied. In this situation, only mixed strategy equilibrium exists: each unemployed worker searches with \( \sigma \in (0, 1) \) such that everyone is indifferent between searching and not searching. That is,

\[
a[\max\{U^N, U^S\} - (u + \delta U)] - c_S = 0 \quad (23)
\]

It can be readily checked that there is a unique \( a \) (hence a unique \( \sigma \)) satisfying (23).

The new programming problem for each firm is as follows, taking the job acquisition rate \( a \) as given:

\[
\begin{align*}
\min_{w, b} & \{w + b\} \\
\text{subject to} & \\
U^N & \geq \overline{U} \quad \text{(IRW1)} \\
a[U^N - (u + \delta U)] - c_S & \geq 0 \quad \text{(IRW2)} \\
V^N & \geq \overline{V} \quad \text{(IRF)} \\
U^N & \geq U^S \quad \text{(ICW)} \\
V^N & \geq V^C \quad \text{(ICF)}
\end{align*}
\]

Compared to the programming problem in the basic model, condition IRW2 is added because firms have to induce workers to search. Given the IRW2 condition and the ICW condition, the IRW1
condition can be rewritten as
\[ U^N - (u + \delta U) \geq \frac{-c_S}{1 - a} \]

The IRW2 condition can be simplified as
\[ U^N - (u + \delta U) \geq \frac{c_S}{a} \]

Thus the IRW1 condition is redundant. After some algebra, the programming problem can be simplified as the following.

\[
\begin{align*}
\min_{w, b} \{w + b\} \\
\text{subject to:} \\
w + b - v &\geq u + \frac{c_S}{a}(1 - \delta \rho) \quad \text{(IRW2)} \\
w &\geq u + \frac{v - b - \delta \rho c_S}{(1 - a)\delta \rho} \quad \text{(ICW)} \\
b &\leq \delta \rho c \quad \text{(ICF)}
\end{align*}
\]

As in the basic model, the optimal bonus is \( b^* = \delta \rho c \), the highest one that is allowed by the ICF condition. Depending on parameter values, we have the following two scenarios.

**Efficiency wages are not necessary** This scenario arises if \( \delta \rho (c + c_S) \geq v \). The ICW condition is redundant if the IRW2 condition is satisfied. In the optimal contract, IRW2 is binding with \( a = 1 \), that is,
\[ w^* + b^* = v + u + c_S(1 - \delta \rho) \]

This is because under optimal contracts unemployed workers will adjust their search behavior such that the job acquisition rate \( a = 1 \). More precisely, under the optimal contracts both conditions (C1) and (C2) are violated, therefore unemployed workers must search with some \( \sigma \in (0, 1) \). Moreover, the indifference condition (23) uniquely pins down that \( a = 1 \). The term \( c_S(1 - \delta \rho) \) is the average search costs that a worker expects to incur in each period on the equilibrium path. Obviously, the average labor cost \( (ALC) \) curve is independent of the employment level \( E \), thus is horizontal. More
precisely,

\[ ALC(E) = v + u + (1 - \delta \rho)(c + c_S) \]  \hspace{1cm} (24)

**Efficiency wages are necessary** This scenario arises when \( \delta \rho(c + c_S) < v \). Now the IRW2 condition no longer implies the ICW condition. However, the ICW condition is more stringent than the IRW2 condition when \( a \) goes to 1, and vice versa when \( a \) goes to 0. Therefore, there is an \( \hat{a} \) at which the IRW2 condition and the ICW condition are equally stringent:

\[ u + v - b^* + \frac{c_S}{\hat{a}}(1 - \delta \rho) = u + \frac{v - b^* - \delta \rho c_S}{(1 - \hat{a})\delta \rho} \]  \hspace{1cm} (25)

Note that \( \hat{a} \) is independent of \( E \). When \( a > \hat{a} \), the ICW condition implies the IRW2 condition, thus the ICW condition is binding in the optimal contract. On the other hand, when \( a < \hat{a} \), the IRW2 condition implies the ICW condition, thus the IRW2 condition is binding in the optimal contract.

Observe that when \( a \geq \hat{a} \), firms’ main concern is the ICW condition. As a result they will induce unemployed workers to search with probability 1, thus the effective \( a \) equals to the low bound \( \underline{a} \). However, when \( a < \hat{a} \), firms only need to worry about the IRW2 condition, and they will just pay a wage such that the actual job acquisition rate is \( \hat{a} \), that is,

\[ w^* = u + v - b^* + \frac{c_S}{\hat{a}}(1 - \delta \rho) \]  \hspace{1cm} (26)

To see this, define \( \hat{E} \) such that

\[ \frac{(1 - \rho)\hat{E}}{(L - \rho\hat{E})} \equiv \hat{a} \]

When \( E = \hat{E} \), \( w^* \) defined in (26) satisfies both IRW2 and ICW conditions, and every unemployed worker searches with \( \sigma = 1 \) so the actual job acquisition rate is \( \hat{a} \). When \( E < \hat{E} \), \( a < \hat{a} \). Now given \( w^* \) condition (C1) is not satisfied thus \( \sigma < 1 \). Each unemployed worker will search with the probability such that the actual job acquisition rate \( a \) satisfies the indifference condition (23). It can readily checked that the job acquisition rate satisfying this condition is \( \hat{a} \). This is simply because at \( w^* \), the IRW2 condition is binding at \( \hat{a} \).

Intuitively, with the presence of search costs wage contract needs to serve two purposes: motivate employed workers and induce unemployed workers to search. Both of them require wage to be high
relative to unemployment benefit. And both of them are affected by employment level. When employment level is high, the resulting high job acquisition rate makes motivating workers relatively more difficult. In this case the wage should be set just enough to motivate workers. On the other hand, when employment level is low, inducing search is relatively more difficult. However, because unemployed workers will endogenously adjust their search behavior, wage is set just enough to induce the job acquisition rate $\hat{a}$, at which both the ICW and the IRW2 conditions are binding.

In summary, the optimal contract depends on $E$. If $E \geq \hat{E}$, then the ICW condition is binding,

$$w^* = u + \frac{v - \delta \rho (c + c_S)}{(1 - a) \delta \rho} \text{ and } a = a$$

The corresponding $ALC$ is:

$$ALC(E) = w^* + b^* + (1 - \delta \rho)c = u + \frac{v}{(1 - a(E)) \delta \rho} - \frac{a(E)c + c_S}{1 - a(E)}$$  \hspace{1cm} (27)

On the other hand, if $E < \hat{E}$, then

$$w^* = u + v - \delta \rho c + \frac{c_S}{\hat{a}(1 - \delta \rho)} \text{ and } a = \hat{a}$$

Correspondingly,

$$ALC(E) = u + v + (1 - \delta \rho)c + \frac{c_S}{\hat{a}}(1 - \delta \rho)$$  \hspace{1cm} (28)

![Figure 4: The ALC Curve](image)

The above analysis shows that the ALC curve has two portions with different properties, depending on whether $E$ is bigger or smaller than $\hat{E}$. When $E > \hat{E}$, the ALC curve is upward sloping since
\( \omega \) is increasing in \( E \) (see equation (27)). When \( E < \hat{E} \), the ALC is independent of \( E \), thus the ALC curve is horizontal (see equation (28)). Figure 4 shows the shape of an ALC curve. We summarize these results in the following proposition.

**Proposition 5** If \( \delta \rho (c + cs) \geq v \), the ALC curve is horizontal. If \( \delta \rho (c + cs) < v \), the optimal contract has the following form: 

\[
\begin{align*}
\hat{b} &= \delta \rho c, \\
\hat{w} &= u + \hat{b} - v + \frac{cs}{\alpha} (1 - \delta \rho) \quad \text{if } E \leq \hat{E}, \\
\hat{w} &= u + \frac{v - \hat{b} - \delta \rho cs}{(1 - \sigma) \delta \rho} \quad \text{if } E > \hat{E};
\end{align*}
\]

the ALC curve is horizontal when \( E \leq \hat{E} \), and is upward sloping when \( E > \hat{E} \).

Compared to the basic model, the presence of search costs qualitatively changes the shape of the ALC curve. Consider the scenario where efficiency wages are necessary. Without search costs, ALC is always increasing in employment level. However, with the presence of search costs, the ALC curve is horizontal when the employment is below some cutoff level. Note that ALC is the total wage payment plus a constant term independent of the employment level, so the wage-employment relationship has the same shape as the ALC curve. This is a potentially important result for macroeconomics: the wage-employment relationship changes over the course of business cycle. Specifically, wages are rigid during recessions, and are positively correlated with employment level during booming times. Two main existing theories of wage determination both predict a monotonic relationship between wages and unemployment. The efficiency wage model of Shapiro and Stiglitz shows that there is always a negative relationship between wages and unemployment, while the migration model of Harris and Todaro (1970) predicts that there is always a positive relationship. To the best of our knowledge, our model is the first one theoretically establishing that wage-unemployment relationship might be different in different courses of business cycle.

If unemployed workers only play pure strategies regarding to search, then the total wage payment actually increases as \( E \) decreases further from \( \hat{E} \). This is because when the IR condition is binding, decreases in \( E \) directly reduce the return of search. To induce every unemployed worker to search, total wage payment has to increase. As a result, both the wage curve and the ALC curve has a U shape. This is a more dramatic result: wages are procyclical during booming times and countercyclical during recessions. In reality, we believe that unemployed workers are able to adjust their search behavior, but not perfectly as our formal model assumed (because it needs perfect coordination).
Therefore, more realistically wages are slightly countercyclical during recessions.

**Comparative statics** From equations (24), (27) and (28), an increase in $c$ results in a higher ALC if efficiency wages are not necessary, and reduces ALC if efficiency wages are necessary. These properties are the same as those in the basic model. However, with the presence of search costs, a change in $c$ affects $\hat{E}$, the kink of the ALC curve. Consider the situation that efficiency wages are necessary. From equation (25), an increase in $c$ results in a higher $b^*$, thus $\hat{a}$ (hence $\hat{E}$) increases. Therefore, an increase in $c$ would expand the range where the ALC is horizontal. Intuitively, an increase in $c$ reduces the amount of efficiency wages that are necessary to motivate workers, thus inducing workers to search becomes relatively more difficult. As a result, the job acquisition rate $\hat{a}$ at which both the ICW and IRW2 conditions are binding increases.

![Figure 5: The Impact of Changes in $c$ on the ALC Curve](image)

Figure 5 depicts how an increase in $c$ affects the ALC curve. Two labor markets are the same in other parameter values except that $c < c'$. In both markets efficiency wages are necessary. In the figure, $\text{ALC}(c')$ always lies below $\text{ALC}(c)$. And the kink $\hat{E}' > \hat{E}$.

How does an increase in $c_S$ affect the ALC curve? If initially $\delta \rho(c + c_S) \geq v$ (no efficiency wages are necessary), then $\hat{a}$ always equals to 1. From (24), an increase in $c_S$ shifts the whole ALC curve up. This effect is the same as an increase in $c$ in the basic model: above some level, more friction in labor markets only increases the average labor cost.
When \( \delta \rho (c + c_S) < v \), the impact of an increase in \( c_S \) on ALC is subtle. From (25), \( \hat{a} \) and \( \hat{E} \) increases as \( c_S \) increases. This is because an increase in \( c_S \) makes motivating workers easier and inducing workers to search more difficult. For \( E > \hat{E}' \) (where \( \hat{E}' \) is the new cutoff value under \( c'_S > c_S \)), equation (27) shows that \( \text{ALC}(c'_S) < \text{ALC}(c_S) \). This is because an increase in \( c_S \) decreases the necessary wage premium. For \( E < \hat{E} \), it seems ambiguous whether ALC increases or decreases, since an increase in \( c_S \) is accompanied by an increase in \( \hat{a} \). However, we shall show in lemma 2 that these two effects exactly offset each other, so ALC is unaffected by changes in \( c_S \).

**Lemma 2** When \( \delta \rho (c + c_S) < v \), \( \frac{c_S}{\hat{a}} \) remains constant as \( c_S \) changes.

**Proof.** See Appendix. □

![Figure 6: The Impact of Changes in \( c_S \) on the ALC Curve](image)

Figure 6 shows how an increase in \( c_S \) affects the ALC curve. Two labor markets are the same in other parameter values except that \( c_S < c'_S \). In both markets efficiency wages are necessary. In the figure, the kink \( \hat{E}' > \hat{E} \). For \( E \leq \hat{E} \), \( \text{ALC}(c_S) \) equals to \( \text{ALC}(c'_S) \); and for \( E > \hat{E} \), \( \text{ALC}(c'_S) \) always lies below \( \text{ALC}(c_S) \). We summarizes the comparative statics results in the following proposition.

**Proposition 6** When efficiency wages are not necessary, increases in \( c \) or \( c_S \) always shift the ALC curve up. When efficiency wages are necessary, an increase in \( c \) shifts the whole ALC curve downwards, and the domain of \( E \) such that the ALC curve is horizontal expands; an increase in \( c_S \) does not change the horizontal portion of the ALC curve, but its domain expands, and it shifts down the increasing portion of the ALC curve.

32
Proposition 6 has two empirical implications. First, an increase in either $c$ or $c_S$ causes the wage rigidity region to expand, and the wage-procyclical region becomes smaller. Second, an increase in $c$ or $c_S$ makes wages in the wage-procyclical region less sensitive to employment levels; this is simply because the required wage premium is smaller.

6 Empirical Implications and Evidence

Table 1 and 2 summarize the empirical predictions of the basic model and the extended model respectively. Note that all these predictions are essentially comparative statics results.

<table>
<thead>
<tr>
<th>Table 1: Predictions of the Basic Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low turnover costs</td>
</tr>
<tr>
<td>Contract Form</td>
</tr>
<tr>
<td>Total Wage Payment</td>
</tr>
<tr>
<td>Unemployment</td>
</tr>
<tr>
<td>Wage-unemployment Elasticity</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>Table 2: Predictions of the Extended Model</th>
</tr>
</thead>
<tbody>
<tr>
<td>Low turnover costs</td>
</tr>
<tr>
<td>The Wage Curve</td>
</tr>
<tr>
<td>The Rigid Wage Region</td>
</tr>
<tr>
<td>Wage-unemployment Elasticity</td>
</tr>
</tbody>
</table>

Forms of contracts There is little empirical work on the relationship between occupations and bonuses. The main difficulty of doing such empirical work is that most data sets do not distinguish between bonuses and contractually set incentive pay. For data sets that do distinguish the two only the frequency of bonuses are reported, but not the amount of bonuses. Note that the prediction of our model is about different amount of merit pay in different occupations. To proceed, we simply make the assumption that the frequency of bonuses and the amount of bonuses are positively correlated.

$^{17}$This result holds only if turnover costs are not too high.
Table 3 is excerpted from table 3B of MacLeod and Parent (1998), based on the NLSY data (1988-1990). From the table, managers have the highest bonus payment; food and cleaning service workers have the lowest bonus payment; and professionals and secretaries are paid with medium amount of bonus. Similar pattern holds when bonus plus promotion is used as the measure of discretionary pay. This pattern is largely consistent with the predictions of our model. This is because managers usually have high turnover costs, service workers have low turnover costs, and professionals and secretaries have medium turnover costs.

The turnover costs for managers usually are high for two reasons. First, some firm-specific knowledge is needed for a manager to be effective in a firm, and it takes time for a new manager to acquire this knowledge. Second, each managing job may require a different combination of skills and personalities, so finding appropriate candidates for a vacancy takes long time and requires substantial effort. As a result, the recruiting costs for managers are relatively high. On the other hand, the jobs for food service and cleaning workers are fairly standard across firms. Therefore, their turnover costs are usually low. It is also reasonable to think that the turnover costs for professionals and secretaries are lower than those of managers and higher than those of service workers, since their jobs usually have a firm-specific component which is smaller than managers’ but bigger than service workers’.

<table>
<thead>
<tr>
<th>Occupations</th>
<th>Bonus</th>
<th>Bonus+Promotion</th>
</tr>
</thead>
<tbody>
<tr>
<td>Managers</td>
<td>28.46%</td>
<td>47.37%</td>
</tr>
<tr>
<td>Professionals</td>
<td>15.46%</td>
<td>29.22%</td>
</tr>
<tr>
<td>Secretaries</td>
<td>11.60%</td>
<td>25.60%</td>
</tr>
<tr>
<td>Food service workers</td>
<td>7.49%</td>
<td>18.66%</td>
</tr>
<tr>
<td>Cleaning service workers</td>
<td>7.43%</td>
<td>17.33%</td>
</tr>
</tbody>
</table>

Similar pattern emerges from the 1990 British Workplace Industrial Relations Survey (WIRS), which is reported in table 4. Manual workers have the smallest bonus while managers have the highest bonus. Moreover, among manual workers the incidence of bonuses are decreasing in their skills. This is also largely consistent with our model, since turnover costs in an occupation are
roughly increasing in required skills: the more skills a job requires, the more firm-specific skills are involved, hence the higher the turnover costs.

<table>
<thead>
<tr>
<th>Occupations</th>
<th>Incidence of Bonus</th>
</tr>
</thead>
<tbody>
<tr>
<td>Professional and Managerial</td>
<td>35%</td>
</tr>
<tr>
<td>Supervisors</td>
<td>32%</td>
</tr>
<tr>
<td>Clerical, administrative, and secretarial</td>
<td>30%</td>
</tr>
<tr>
<td>Skilled Manual</td>
<td>22%</td>
</tr>
<tr>
<td>Semiskilled Manual</td>
<td>16%</td>
</tr>
<tr>
<td>Unskilled Manual</td>
<td>11%</td>
</tr>
</tbody>
</table>

Table 4: Occupations and Bonuses

Though these empirical evidence are largely consistent with our predictions, they are not a test of our theory. We hope that some carefully designed empirical work can be done in the near future to directly test our model.

**Wage differentials** Using the data of Industrial Wage Survey, Brown (1992) conducts an empirical study on the relationship between wage levels and methods of pay. He finds that the workers paid by standard rates on average earn a higher wage than the workers with merit pay. Standard rates and merit pay correspond to efficiency wages and subjective performance pay respectively in theoretical models. This empirical evidence is consistent with the prediction of our model: merit pay can reduce the amount of efficiency wages, so workers paid by standard rates enjoy a higher wage premium, hence earn more than workers with merit pay.

**Methods of pay and unemployment** One implication of our model is that the level of equilibrium unemployment is an decreasing function of the usage of bonuses. Based on the 1990 British WIRS data, MM find that there is a negative correlation between the percentages of workers with merit pay and unemployment rates among occupations. The study by MacLeod and Parent (1998) further supports this result. Using the data of NLSY 1988-1990, they show that there is a strong negative relationship between the use of discretionary bonus and the local unemployment.
**The wage curve** Using aggregate time series, early empirical literature found that real wages are acyclical during business cycles. Recently, Solan, Barsky and Parker (1994) argued that “the true procyclicality of real wages is obscured in aggregate time series because of a composition bias.” Using longitudinal micro data, they find that real wages in the US have been substantially procyclical since the 1960s. However, like most of the empirical works in this literature, they just test whether real wage is procyclical but do not estimate the whole wage curve.

Fortunately, a small but important literature initiated by Blanchflower and Oswald (1994, BO hereafter) does estimate the whole wage curve. Using the US General Social Survey (GSS) data (1974-1988), they estimate industry wage curve (wage as a function of unemployment rate in industries) and regional wage curve (wage as a function of regional unemployment rate). Figure 7, copied from BO (page 107), illustrates their estimation result. Both curves are initially downward-sloping and then become upward-sloping. Both wages are minimized at an unemployment rate of approximately 6%-8%. These wage curves are consistent with the empirical predictions of our extended model. The upward-sloping portion of the wage curve suggests that unemployed workers are not able to adjust perfectly their search behavior.

![Figure 7: An Estimated Wage Curve](image)

BO also estimate the wage curve based on the US Current Population Surveys (CPS, 1964-1991). The results are slightly different from those from GSS: the wage curve is significantly downward
sloping when unemployment rate is low, and it flattens out as unemployment rate increases, but there is no upward-sloping portion of the wage curve. Similar estimation result holds for the data of the British Social Attitude Surveys (1983-1989): the wage curve flattens out when unemployment rate is greater than 13%. Using the General Household Surveys data (1973-1977) of Britain, BO find that the wage curve has a U shape, with the turning point occurring around unemployment rate 4.5%. Based on the International Social Survey Program data (1986-1991) of West Germany, the estimation of BO shows that the wage curve flattens out around unemployment rate 11%. Bratsberg and Turunen (1996) estimate the US wage curve for young workers using the 1979-1993 waves of NLSY. According to their study, the wage curve based on annual earnings flattens out when unemployment rate is higher than 12%, and the wage curve based on hourly wage exhibits U shape, with the minimum wage reached at 11.5% unemployment rate.

Though there are some minor differences, all the above estimation results show that the wage curve either flattens out or becomes upward-sloping at fairly high unemployment rate. They are largely consistent with our empirical prediction that wages are procyclical during booms and either rigid or counter-cyclical in recessions.

**Wage-unemployment elasticity**  BO also estimate the wage-unemployment relationship for different occupations using the data of 1990 British WIRS. The results are reported in table 5. Unskilled manual workers have the largest wage-unemployment elasticity, supervisors have the lowest elasticity, and clerical workers have some medium elasticity. As we argued before, turnover costs are increasing in the order of unskilled manual workers, clerical workers and supervisors. Therefore, the pattern of wage-unemployment elasticity is largely consistent with the predictions of our model. Unskilled manual workers have the lowest turnover costs, hence their methods of pay are mainly efficiency wages. As a result, their wages are more procyclical. On the other hand, supervisors (managers) have high turnover costs, thus their methods of pay are mainly subjective performance pay, which lead to low wage-unemployment elasticity.
Table 5: Occupations and Wage-unemployment Elasticity

<table>
<thead>
<tr>
<th>Occupations</th>
<th>Unskilled manual</th>
<th>Skilled manual</th>
<th>Clerical</th>
<th>Supervisors</th>
</tr>
</thead>
<tbody>
<tr>
<td>Coefficient</td>
<td>-.0916</td>
<td>-.0325</td>
<td>-.0434</td>
<td>-.0048</td>
</tr>
</tbody>
</table>

Table 6 reports the results of two other studies. Again managers have the lowest wage-unemployment elasticity, and manual workers or Clerks have the highest elasticity.

Table 6: Occupations and Wage-unemployment Elasticity

<table>
<thead>
<tr>
<th>Occupations</th>
<th>Manual</th>
<th>Clerks</th>
<th>Professionals</th>
<th>Managers</th>
</tr>
</thead>
<tbody>
<tr>
<td>British GHS data</td>
<td>-.0721</td>
<td>-.0631</td>
<td>-.0497</td>
<td></td>
</tr>
<tr>
<td>Australian ABS IDS data</td>
<td>-.0896</td>
<td>-.0224</td>
<td>-.0198</td>
<td></td>
</tr>
</tbody>
</table>

7 Conclusions

We studied contract selection between efficiency wages and subjective performance pay to motivate workers in a labor market setting. Though subjective performance pay is “cheaper” than efficiency wages, it is limited by firms’ incentive to renege. The presence of the turnover costs borne by firms reduces firms’ incentives to renege, thus making implicit bonus credible to some extent. In the optimal contracts, the amount of bonus is positively correlated, and the amount of wage premium is negatively correlated, with the turnover costs borne by firms. Up to some threshold, an increase in turnover costs effectively reduces the total wage payment and total labor costs, thus increases the equilibrium employment level and social welfare. This is a surprising result: a little bit friction in markets can actually be beneficial. We conjecture that this result also holds in more general settings where both agents in a relationship have moral hazard problems and both are able to change partners in markets.

The extended model incorporates workers’ search costs. In this setting firms need not only to motivate workers but also to induce unemployed workers to search. Wage-unemployment relationship turns out to be different during booms and recessions: wages are procyclical in booms, and are either rigid or countercyclical in recessions. Our model generates rich empirical implications. The forms of

---

wage contracts and total wage payments are different in different occupations with turnover costs. Occupations using more bonus payments have lower total wage payments and lower unemployment rates. Occupations with high turnover costs have low wage-unemployment elasticity. Some empirical evidence is consist with these predictions.

On the empirical side, more empirical works need to be done to directly test the implications of our model. On the theoretical side, it would be interesting to incorporate workers’ heterogeneity into the model. One of my working papers (Yang, 2004) studies this problem. Another interesting extension would be to model relational contracts in a setting where demand for labor is fluctuating over time. By doing that we are able to see more explicitly how wages and contract forms change during the course of business cycle. This is left for future research.

8 Appendix

Proof of Proposition 2 Proof. The proof is by construction. Define \( w_p \) such that

\[
\bar{U}(w^*, b^*) = (w_p + b^* - v) + \delta [\rho U^N(w^*, b^*) + (1 - \rho) \bar{U}(w^*, b^*)]
\]

and \( w_r \) such that

\[
\bar{V}(w^*, b^*) = (p - w_r - b^*) + \delta \rho V^N(w^*, b^*)
\]

The relational contract specifies three states: cooperative state, punishment state for the worker and punishment state for the firm; in any case the relationship continues. In the cooperative state the payment scheme is \((w^*, b^*)\); in the punishment state for the worker, the payment scheme is \((w_p, b^*)\); in the punishment state for the firm, the payment scheme is \((w_r, b^*)\). The transition of states is the following: the game starts in cooperative state; if last period is in cooperative state and no party deviates, stay in the cooperative state; if last period is in cooperative state and the worker (firm) deviates, transit to the punishment state for the worker (firm); if last period is in punishment state and no party deviates, transit to cooperative state; if last period is in punishment state and the worker (firm) deviates, transit to the punishment state for the worker (firm). Note that \( w_p < w^* \), which punishes the worker if he deviates. By the definition of \( w_p \), it may be less than 0. On the
other hand, \( w_r > w^*_r \), which is designed to punish the firm if it reneges.

In cooperative state, if the worker shirks, his continuation payoff inside the relationship is \( \overline{U} \) by construction, which is the same as his outside option, so the worker is willing to stay in the relationship and accept the punishment. Note that the No-shirking condition is satisfied since \( U^N(w^*, b^*) \geq U^S(w^*, b^*) \). In cooperative state, if the firm reneges, his continuation payoff inside the relationship is \( \overline{V} \) by construction, so he will continue the relationship and accept the punishment; and the No-reneging condition is satisfied since \( V^N(w^*, b^*) \geq V^R(w^*, b^*) \). In punishment states, if the worker deviates, his payoff is \( U^{PS} = w_i + \delta U \) (where \( i = r, p \) depends on the punishment state for the worker or the firm). If the worker doesn’t deviate, his payoff is \( U^P = (w_i + b - v) + \delta [\rho U^N + (1 - \rho)U] \). Simple algebra shows that

\[
U^P \geq U^{PS} \iff \delta \rho (U^N - \overline{U}) \geq (v - b) \iff U^N(w^*, b^*) \geq U^S(w^*, b^*)
\]

which is satisfied. So the worker has no incentive to deviate in punishment states. In punishment states, if the firm reneges, its payoff is \( V^{PR} = (p - w_i) + \delta \rho \overline{V} \); if it pays the promised bonus, its payoff is \( V^P = (p - w_i - b^*) + \delta \rho V^N \). The firm has no incentive to deviate if and only if

\[
V^P \geq V^{PR} \iff b \leq \delta \rho c \iff V^N(w^*, b^*) \geq V^R(w^*, b^*)
\]

which is satisfied. Therefore, the specified strategies constitute a subgame perfect equilibrium. Moreover, the relationship is always continued, so the Pareto frontier is always reached. Therefore, it’s strongly renegotiation proof.

Proof of Lemma 2  Proof. Let \( k \equiv v - b^* \) and \( z \equiv \delta \rho \). Then (25) becomes

\[
k + \frac{cs}{a}(1 - z) = \frac{k - zcs}{(1 - a)z}
\]

(29)

Differentiate (29) we get

\[
\frac{dcS}{da} = \frac{k - zcs}{(1 - a)z} + \frac{cs}{a}(1 - z)(1 - a)
\]

\[
= \frac{k \hat{a} + \frac{cs}{a}(1 - z)}{(1 - z) + z \hat{a}}
\]
It follows that
\[
\frac{dc_S}{da} - \frac{c_S}{a} = \frac{k\hat{a} - zc_S}{(1-z) + za} \tag{30}
\]
Solve \(c_S\) from (29) and substitute it into (30), we get
\[
\frac{dc_S}{da} - \frac{c_S}{a} = \frac{1}{(1-z) + za} \left[ k\hat{a} - \frac{k\hat{a} - kz\hat{a}(1-\hat{a})}{(1-z) + za} \right]
\]
\[
= \frac{1}{(1-z) + za} \left\{ k\hat{a}[(1-z) + za] - k\hat{a} + kz\hat{a}(1-\hat{a}) \right\}
\]
\[
= 0
\]
Therefore, \(\frac{dc_S}{da}\) caused by a change in \(c_S\) always equals to \(\frac{c_S}{a}\). This implies that \(\frac{c_S}{a}\) does not change as \(c_S\) changes. It follows immediately from (28) that for \(E < \hat{E}\), ALC remains the same as \(c_S\) changes.

\textbf{References}


