EFFICIENCY WAGES AND SUBJECTIVE PERFORMANCE PAY

HUANXING YANG*

This paper studies optimal relational contracts in motivating workers in a market setting. We find that labor markets with higher turnover costs will use more subjective performance pay and less efficiency wages and that in those markets, the total wage payment is lower and the equilibrium employment level is higher. Surprisingly, under certain conditions, an increase in turnover costs leads to higher social welfare. Incorporating workers' search costs, we show that wages are procyclical in booms and are either rigid or countercyclical during recessions. The predictions of the model are consistent with some empirical evidence. (*JEL* D82, J33, J41, J63)

I. INTRODUCTION

How to design incentive schemes to motivate workers is an important topic in economics. The shirking models of efficiency wages, such as Shapiro and Stiglitz (1984), establish that firms need to pay a wage premium (efficiency wages) to motivate workers, with unemployment serving as a punishment device. However, one shortcoming of these models is that performance pay plays no role. One justification for their omission of performance pay is that individual performance may not be verifiable. Nevertheless, if workers' performance is observable and employment relationships are repeated, firms can use implicit bonuses or relational contracts based on workers' subjectively assessed performance to motivate workers. Since subjective performance pay cannot be legally enforced, it has to be self-enforcing.

Given that both efficiency wages and subjective performance pay motivate workers, what is the optimal wage contract from the firm's 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? To answer these questions, this paper provides a theory of *contract selection* in a *market* setting.

In a seminal paper, MacLeod and Malcomson (1998, MM hereafter) provided a model of contract selection between efficiency wages and subjective performance pay. The driving force in their model is the 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, no subjective performance pay is credible, and firms have to use solely efficiency wages 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 solely subjective performance pay to motivate workers.

In MM, there are no exogenous turnover costs. Complementary to MM, this paper provides a model of contract selection driven by exogenous turnover costs in labor markets. We focus on the situation in which unemployment always exists in labor markets, thus ruling out market condition as a determinant of

*I wish to thank two anonymous referees, Ken Burdett, Bill Dupor, Jan Eeckhout, Paul Evans, Rafael Rob, and the participants at the Fall 2003 Midwest Theory Conference for helpful comments and discussions. I am truly indebted to Steven Matthews for his invaluable advice and guidance. The usual disclaimers apply.

Yang: Assistant Professor of Economics, Ohio State University, 405 Arps Hall, 1945 N. High Street, Columbus, OH 43210. Phone 614-292-6523, Fax 614-292-3906, Email yang.1041@osu.edu

ABBREVIATIONS

ALC: Average Labor Cost
BO: Blanchflower and Oswald
GSS: General Social Survey
MM: MacLeod and Malcomson
NLSY: National Longitudinal Surveys of Youth
WIRS: Workplace Industrial Relations Survey

Economic Inquiry
(ISSN 0095-2583)
Vol. 46, No. 2, April 2008, 179–196

doi:10.1111/j.1465-7295.2007.00069.x
Online Early publication July 28, 2007
© 2007 Western Economic Association International

contract selection. It turns out that turnover costs affect the amount of efficiency wages and performance pay in optimal contracts. While in MM efficiency wages and subjective performance pay cannot be used together to motivate workers, in our model firms are able to use combinations of both methods of payments. By affecting optimal contracts, turnover costs also have impacts on the equilibrium employment level and social welfare. Finally, our model generates rich empirical implications about the relationships among turnover costs, forms of employment contracts, and levels of employment.

Our basic model studies how turnover costs borne by firms affect contract selection. From the firm's point of view, subjective performance pay is "cheaper" since efficiency wages entail a wage premium. Thus, the optimal wage contract uses the maximum amount of bonus to motivate workers. However, subjective performance pay may not be credible due to the firm's moral hazard problem: in labor markets with positive unemployment, a firm can immediately hire a new worker after reneging on the implicit bonus. The presence of turnover costs can alleviate this moral hazard problem. This is because a worker can punish a reneging firm by quitting and the firm has to incur turnover costs in hiring new workers. Therefore, in the optimal contract, the amount of bonus is increasing in turnover costs. Since subjective performance pay is cheaper, as the turnover costs increase, more subjective performance pay leads to a lower total wage payment.

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 the average labor cost (ALC), which consists of the 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 reduces the ALC and leads to an increase in the equilibrium employment level. 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 of (exogenous) friction in markets is beneficial for social welfare. The main reason behind this result is that friction in markets alleviates the firms' moral hazard problem and gives them commitment power, which in turn grants firms more flexibility to alleviate the workers' moral hazard problem. A general interpretation of this result is that exogenous friction might be more efficient than endogenously created friction (efficiency wages) in overcoming double moral hazard problems in markets.

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). It turns out that inducing workers to search becomes more difficult as the unemployment rate increases. The main result in the extended model is that wages are increasing in the employment level only when employment is high and are completely rigid when employment is low. This implies that wage-unemployment relationship changes over the course of business cycles: wages are procyclical in booms and rigid during recessions.

Our model generates rich empirical implications. First, different labor markets (occupations) will adopt different forms of wage contracts. In particular, the efficiency wage component (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 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 in labor markets. All these predictions are consistent with some empirical evidence.[1]

Relational contracts have been studied by Bull (1987), MM (1989, 1998), Baker, Gibbons, and Murphy (1994), and Levin (2003). However, except for MM (1989, 1998), all the other papers study the 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 contract selection between efficiency wages and subjective performance pay. MM (1989) offered a complete

1. See Section VI for details.

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.

Akerlof and Katz (1989) incorporated a performance bond into a shirking model of efficiency wages. They, however, assumed that firms are able to commit: firms never forfeit a worker's bond if he does not shirk. In contrast to their assumption, our model, following the literature of relational contracts, assumes that firms are not able to commit. The labor turnover models of efficiency wages, such as Salop (1979) and 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 total turnover costs. This result is in direct contrast to the prediction of our model.

The structure of the paper is as follows. Section II sets up the basic model. Section III studies the optimal wage contracts. In Section IV, first the stationary market equilibrium is derived and then comparative statics and welfare analysis are conducted. In Section V, we extend the basic model to incorporate search costs. Section VI presents some empirical evidence. Section VII concludes the paper.

II. THE BASIC MODEL

Consider a labor market over time, with time t being discrete. There are L workers and many firms which create J job vacancies in total. While L is exogenously given, there is free entry on the firms' side, thus J is endogenous. Both workers and firms are risk neutral and share the same discount factor δ. Each job has the same revenue product of labor p. Each firm takes p as given, but in the aggregate, p is a decreasing function of the total employment E. 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 match breaks up with probability $1 - \rho$ for exogenous reasons. 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 $W_t - ve_t$, where W_t is the total wage payment, $v > 0$ is the disutility of work, and e_t is effort in period t. For simplicity, we assume that a worker can either work ($e_t = 1$) or shirk ($e_t = 0$). The profit of a filled job vacancy in period t is $pe_t - W_t$. Workers without a job receive an unemployment benefit $u > 0$ per period, which is exogenously given. Consistent with employment at will, we assume that only spot contracts are legally enforceable. Following the incomplete contract literature, we assume that e_t is observable but not verifiable.[2] 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 bonuses. In particular, W_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 an employed worker shirks, then his employer's period profit is negative. Thus, firms have to motivate workers to exert effort. To make sure there is positive employment, we assume that $p(0) > u + v/(\delta\rho)$. In addition, we assume that $p(L) \leq u + v/(\delta\rho)$. This assumption ensures that $J \leq L$; that is, there is always unemployment and 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.[3] Workers incur search costs in finding jobs. Moreover, each firm may require a skill which is firm specific.[4] 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 $c_F \geq 0$.[5] The parameter c_F captures the

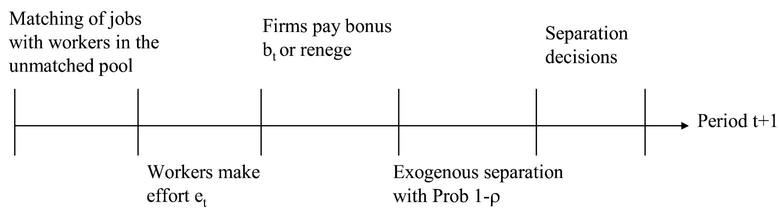
2. An alternative setting would be that, while e_t is not observable, the performance is observable but not verifiable, and e_t determines the worker's performance with some noise. The only difference this alternative setting makes is that shirking cannot be detected for sure. Since workers are risk neutral, the qualitative results of the paper still hold under this alternative setting.

3. For some jobs, Lang (1991) found that firms are willing to pay employment agencies large sums for finding workers.

4. See Hashimoto and Yu (1980) for the issue of firm-specific human capital.

5. c_F can be interpreted as the training cost to acquire firm-specific skills.

FIGURE 1
Timing



degree of firm specificity of jobs: the more firm specific are jobs in the market, 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 is borne by firms alone.[6] Later on, we will discuss how relaxing this assumption affects the results of the model. In the basic model, we ignore workers' search costs, which will be studied in Section V. 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 homogeneous regardless of their employment history.

This is essentially an infinitely repeated game with some employment relationships reshuffled in each period. The timing of a typical period is shown in Figure 1. At the beginning of period t, 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 he rejects the offer, the worker is unemployed for that period. Then, each employed worker chooses effort and gets paid. After the payments have been made, a $1 - \rho$ fraction of existing matches exogenously break up. Then, in each surviving match, the worker and the firm simultaneously decide whether to continue the relationship into the next period.[7] All the unmatched agents enter into the unmatched pool in the beginning of the next period.

The model is similar to MM but with two departures. First, we introduce turnover costs in labor markets, which are not studied in their model. Second, we only consider the case $J \leq L$, while in their model both cases $J \leq L$ and $J > L$ are studied.[8] Note that there are double moral hazard problems: workers may want to shirk and collect the fixed wage w_t, and firms may want to renege on the bonus b_t after workers exert effort. The concern for external reputation may mitigate the moral hazard problems. We rule out such a reputation effect.[9]

We are interested in equilibria in which employed workers exert effort and firms do not renege in each period. There are many relational contracts that can be supported as perfect equilibria, some of which may involve complicated strategies. For simplicity, we restrict our attention to trigger strategies. In particular, if a worker shirks, the firm will fire him immediately. Similarly, if a firm reneges on a bonus (pays less than b_t), the worker quits immediately.[10]

III. THE OPTIMAL CONTRACTS

We are interested in stationary equilibrium, in which the 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 nonshirker (shirker) and $\overline{U}$ as

6. This assumption is not unrealistic. An empirical study by Green et al. (2000) shows that even expenses on most transferable training are paid by employers.

7. Alternatively, it is also possible that endogenous separation decisions are made before exogenous separation occurs. But whether exogenous separation occurs before or after endogenous separation decisions are made does not matter, since a match breaks up if either separation occurs.

8. Another minor difference is that in MM, there is a job creation cost. Incorporating a job creation cost would not change the qualitative results of our paper.

9. External reputation may not work for two reasons. First, outsiders do not know exactly whether a separation occurs due to exogenous reasons or cheating. Second, even if a separation is known to be due to cheating, outsiders do not know exactly which party is at fault.

10. The trigger strategies are appealing in practice: if at least one party violates the relational contract, the informal relationship sours and is unlikely to continue.

the expected discounted lifetime utility of an unemployed worker. Let E be the number of employed workers and a be the job acquisition rate. Note that both E and a are endogenous variables. The value function U^N is the following (supposing firms do not renege):

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

The value of U^N consists of two terms. The first term in Equation (1) is a nonshirker's current period payoff. With probability $1-\rho$, the current match breaks up exogenously so the worker gets a continuation value $\overline{U}(w,b)$. With probability ρ, the current match continues in the next period and the worker's continuation value is $U^N(w,b)$. Similarly, other value functions are:

$$(2)\quad U^S(w,b) = w + \delta\overline{U}(w,b)$$

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

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

Similarly, denote V^N (V^R) as the expected discounted lifetime profit of a filled job vacancy if the firm does not renege (reneges)[11] and $\overline{V}$ as the expected discounted lifetime profit of an unfilled job vacancy. The value functions are the following (supposing workers do not shirk):

$$(4)\quad V^N(w,b) = (p-w-b) + \delta[\rho V^N(w,b) + (1-\rho)\overline{V}(w,b)]$$

$$(5)\quad V^R(w,b) = (p-w) + \delta\overline{V}(w,b)$$

$$(6)\quad \overline{V}(w,b) = V^N - c.$$

11. Given that workers are playing trigger strategy, for a reneging firm, paying 0 instead of any positive bonus less than b is the most profitable deviation.

The continuation value of V^N has two components. With probability $1-\rho$, exogenous separation occurs, so the firm gets a continuation value $\overline{V}$; with probability ρ, the existing match remains in the next period, and the firm's continuation value is $V^N(w, b)$. V^R is different from V^N in continuation values: after the firm reneges, the existing match breaks for sure so the firm's continuation value is always $\overline{V}$. $\overline{V}$ is the firm's fallback position if it reneges. Since there is unemployment in the market, the firm can immediately find a new worker to fill its vacancy, by incurring turnover cost c. Therefore, $\overline{V}(w,b) = V^N - c$.

A. 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 earn zero profit due to free entry, 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:

$$\min_{w\geq 0, b\geq 0} W = w+b$$

subject to

$$U^N \geq \overline{U} \quad \text{(IRW)}$$
$$\overline{V} = 0 \quad \text{(IRF)}$$
$$U^N \geq U^S \quad \text{(ICW)}$$
$$V^N \geq V^R \quad \text{(ICF)}.$$

The IRF condition depends on the revenue product of labor $p(E)$. For the time being, we ignore this condition and discuss it in the next section when we study market equilibria. Substituting Equations (1) and (2), the no-shirking condition (ICW) becomes:

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

The left-hand side of Equation (7) is the rent of continued employment enjoyed by an employed 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 firms can also use a combination of both.

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

$$b \leq \delta\rho c. \tag{8}$$

The left-hand side of Equation (8) is the firm's current period gain from reneging; the right-hand side is the expected cost of reneging. 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.[12] Therefore, any credible b has an upper bound $\delta\rho c$.

B. Optimal Contracts

With a positive c, firms have some freedom to choose b in relational contracts. From the no-reneging condition (Equation 8), the credible bonuses the firm can choose are 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 $c \in [0, v/(\delta\rho))$, in the optimal relational contract, the firm should set $b^* = \delta\rho c$. If $c \geq v/(\delta\rho)$, in the optimal relational contracts, the firm can set any $b^* \in [v, \delta\rho c]$ subject to $w^* = u + v - b^* \geq 0$.*

C. Proof. See the Appendix. ■

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 the firms' perspective. This is because performance pay discourages shirking directly without increasing total compensation, while efficiency wages require firms to increase total compensation. 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 the 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.

Define the wage premium w^p as the extra utility per period enjoyed by an employed worker relative to an unemployed worker. By Lemma 1, we can calculate w^p in the optimal contracts.

$$\begin{aligned} w^p &= (b^* + w^* - v) - u \\ &= \{1/[(1-a)\delta\rho] - 1\}\max\{v - \delta\rho c, 0\} \end{aligned} \tag{9}$$

From Equation (9), it is clear that the wage premium is decreasing in c. Moreover, when $c \geq v/(\delta\rho)$, the wage premium equals to 0. On the other hand, when $c = 0$, $b^* = 0$ and firms have to use solely efficiency wages to motivate workers. Note that this case corresponds to the situation studied by Shapiro and Stiglitz (1984). The following proposition summarizes the above results.

PROPOSITION 1. *(a) Suppose $c \geq v/(\delta\rho)$. Then, 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, and the optimal contract is purely in the form of performance pay. (b) Suppose $c \in (0, v/(\delta\rho))$. Then, the optimal contract is the following: $b^* = \delta\rho c$ and $w^* = u + (v - \delta\rho c)/[(1 - a)\delta\rho]$. Employed workers receive a positive wage premium, which is decreasing in c. The optimal contract is a combination of performance pay and efficiency wages. (c) Suppose $c = 0$. Then, the optimal contract has no performance pay component and is in the form of pure efficiency wages.*

D. Robustness

Note that the trigger strategies associated with the optimal contract do constitute a subgame-perfect Nash equilibrium even when

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 $J \leq L$.

$c \in [0, v/(\delta\rho))$ (so employed workers enjoy a positive wage premium). 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 in separation decision would not change the outcome.[13]

So far, we have assumed that c is borne by firms solely. Now suppose that the firm and the new worker share the turnover cost c according to some bargaining rule, with the firm bearing cost θc 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 $\theta(\delta\rho c)$, which is less than $\delta\rho c$ 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 later. The amount of necessary efficiency wages is still decreasing in c if $c \in [0, v/(\delta\rho))$. Overall, the optimal bonus is decreased, but the efficiency wage is more or less the same. The qualitative results of the model remain the same.

Note that any severance pay has no value in overcoming the moral hazard problems. Suppose employment contracts specify that firms pay workers s whenever separation occurs. This severance pay s enables firms to credibly enforce a bonus s. Under this circumstance, however, the workers' moral hazard problem is not altered: though workers do not receive the bonus s in the case of shirking, they receive a severance pay s instead, thus there is no punishment for shirking.

IV. MARKET EQUILIBRIUM

In stationary market equilibria, all the firms in the market will offer the same optimal contract derived in the previous section since they face the same programming problem. Moreover, in equilibria, all employed workers exert effort, all firms pay the implicit bonus (if there is any), and employment relationships end only due to exogenous separation. Denote E^* as the employment level(s) in stationary equilibria, which is determined by the firms' free entry condition:

$$(10)\quad \overline{V}(w^*, b^*, E^*) = [p(E^*) - w^* - b^* - \delta(1-\rho)c]/(1-\delta) - c = 0 \Leftrightarrow p(E^*) = w^* + b^* + (1 - \delta\rho)c.$$

The equilibrium job acquisition rate a is

$$(11)\quad a = (1 - \rho)E^*/(L - \rho E^*).$$

Using Equation (11) and the results in Proposition 1, Equation (10) can be written more explicitly. Specifically, for $c \in [0, v/(\delta\rho))$, Equation (10) becomes

$$(12)\quad p(E^*) = u + v/(\delta p) + [v/(\delta\rho) - c](1 - \rho)E^*/(L - E^*).$$

For $c \geq v/(\delta\rho)$, Equation (10) can be rewritten as:

$$(13)\quad p(E^*) = u + v/(\delta\rho) + (1 - \delta\rho)[c - v/(\delta\rho)].$$

The right-hand side of Equation (10) specifies the ALC per period on the equilibrium path, with $(1 - \delta\rho)c$ being the average turnover costs incurred per period. Now, define

$$(14)\quad \text{ALC}(E) \equiv w^*(E) + b^*(E) + (1 - \delta\rho)c = \begin{cases} u + v/(\delta\rho) + [v/(\delta\rho) - c][(1-\rho)E^*]/(L - E^*) & \text{if } c \in [0, v/(\delta\rho)) \\ u + v/(\delta\rho) + (1 - \delta\rho)[c - v/(\delta\rho)] & \text{if } c \geq v/(\delta\rho) \end{cases}.$$

Thus, ALC(E) specifies an ALC curve. According to Equations (12) and (13), market equilibria are determined by the intersection(s) of the revenue product curve $P(E)$ and the ALC curve ALC(E). It can be easily seen from (14) that ALC is strictly increasing in E if $c \in [0, v/(\delta\rho))$, and ALC is independent of E if $c \geq v/(\delta\rho)$. On the other hand, the revenue product curve $p(E)$ is downward sloping. Therefore,

13. Note that since any endogenous separation results in welfare loss (turnover costs c have to be incurred), the equilibria derived under trigger strategies are not renegotiation proof. In the previous version, we show that the optimal contract derived under trigger strategies can be supported as the optimal contract in a strongly renegotiation proof equilibrium in the sense of Farrell and Maskin (1989). The proof is available upon request.

the market equilibrium must be unique.[14] Figure 2 illustrates the determination of market equilibrium for the case $c \in [0, v/(\delta\rho))$, where point A denotes the market equilibrium.

A. Comparative Statics

We are interested in how changes in c affect the market equilibrium.

PROPOSITION 2. *For $c \in [0, v/(\delta\rho))$, the equilibrium employment level E^* is increasing in c. For $c \geq v/(\delta\rho)$, E^* is decreasing in c.*

Proof. First consider the case $c \in [0, v/(\delta\rho))$. Differentiating Equation (12) with respect to c, we get

$$(15) \quad \partial E^*/\partial c = (1-\rho)E^*(L-E^*) / \{[v/(\delta\rho) - c](1-\rho)L - p'(E^*)(L-E^*)^2\} > 0,$$

since $E^* \in (0, L)$ and $p'(E^*) < 0$. If $c \geq v/(\delta\rho)$, then by Equation (13)

$$\partial E^*/\partial c = (1-\delta\rho)/p'(E^*) < 0. \quad \blacksquare$$

It is surprising that the equilibrium employment level E^* is increasing in c when c is small. The intuition for this result is the following. An increase in turnover costs by Δc can decrease the total wage payment (or the wage premium) in every period by $\{1/[(1 - a)\delta\rho] - 1\}\delta\rho\Delta c$. On the other hand, the average turnover costs per period on the equilibrium path increase by $(1 - \delta\rho)\Delta c$. Overall, the first effect dominates the second one. Intuitively, an increase in turnover costs reduces wage payment in *each period*, while the increase in average turnover costs per period is small since each job only incurs the turnover costs *occasionally* (with probability $1 - \rho$) on the equilibrium path. Therefore, an increase in c shifts the ALC curve downward, leading to an increase in E^*.

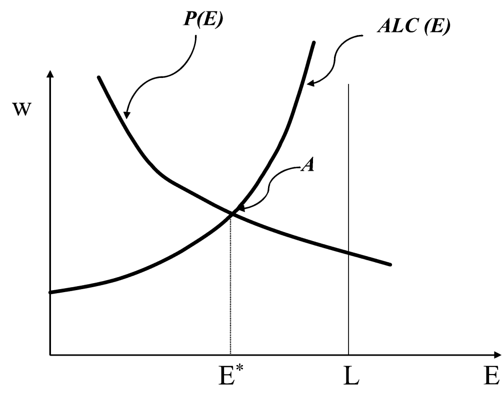
This result is reversed when $c \geq v/(\delta\rho)$. If c falls in this region, by Equation (14), an increase in c shifts the ALC curve upward. This is because an increase in c cannot reduce the wage payment further (the IRW condition is binding), but it directly pushes up the average turnover costs per period. Hence, the equilibrium employment level E^* is decreasing in c. By Equation (14), it can be easily seen that an increase in ρ or δ shifts the ALC curve down. Thus, as the exogenous separation rate $(1 - \rho)$ decreases or the discount factor δ increases, the equilibrium employment level E^* increases.

Changes in c also affect the wage-employment relationship. For $c \in [0, v/(\delta\rho))$,

$$\partial(w^* + b^*)/\partial E = [v/(\delta\rho) - c](1-\rho)L / (L-E)^2.$$

Thus, an increase in c reduces the wage-employment elasticity. This is due to the fact that an increase in c reduces the required wage premium, thus making wages less sensitive to the employment level.

FIGURE 2
Market Equilibrium



B. Welfare Properties

Now, we study how changes in c affect social welfare. In the market equilibrium, each firm's expected profit is 0 since $p(E^*) = \text{ALC}(E^*)$. The total social surplus S (per period) of the market equilibrium with employment E^* is

$$(16) \quad S(E^*, c) = \int_0^{E^*} p(E)\,\mathrm{d}E - E^*(u+v) - (1-\rho)E^*c.$$

14. For $c \in [0, v/(\delta\rho)]$, the existence of equilibrium $E^* \in (0, L)$ is guaranteed if $p(0) > u + v/(\delta\rho)$. For $c \geq v/(\delta\rho)$, the existence of equilibrium $E^* \in (0, L)$ is ensured if $p(L) < u + v/(\delta\rho) + (1 - \delta\rho)[c - v/(\delta\rho)] < p(0)$.

The first term is the social value of the total output or consumers' total willingness to pay for the total output. The second term is the total cost of labor, and the third term is the total turnover cost incurred per period. Taking the derivative of Equation (16) with respect to c,

$$(17) \qquad \partial S/\partial c = [p(E^*) - (u+v) - (1-\rho)c]\partial E^*/\partial c - (1-\rho)E^*.$$

We first consider the case $c \in [0, v/(\delta\rho))$. Substituting $\partial E^*/\partial c$ from Equation (15) into Equation (17), we get

$$(18) \qquad \partial S/\partial c > 0 \Leftrightarrow v(1-\delta)/\delta > -p'(E^*)(L-E^*).$$

Condition (18) 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 17). 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 increasing the social welfare. If the revenue product of labor is elastic enough, a small increase in c can induce a big increase in the equilibrium employment level, causing the positive effect to dominate the negative effect and social welfare increases.[15]

When $c \geq v/(\delta\rho)$, 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 (17). Therefore, we have the following proposition.

PROPOSITION 3. *If $c \in [0, v/(\delta\rho))$ and Condition (18) is satisfied (the revenue product of labor $p(E)$ is elastic enough), the social surplus is increasing in c. When $c \geq v/(\delta\rho)$, the social surplus is decreasing in c.*

Proposition 3 is a surprising result: a little bit of (exogenous) friction in markets can improve social welfare. Conventional 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 of (exogenous) friction in the market can actually 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), a certain amount of matching friction has to be created endogenously by using efficiency wages. The presence of some exogenous friction alleviates the firms' moral hazard problem and gives them commitment power, which makes contingent contracts (performance pay) feasible. Contingent contracts not only reduce the amount of endogenously created friction that is necessary to motivate workers but also reduce the total amount of friction in the market that is necessary to motivate workers. Broadly speaking, this result implies that exogenous friction might be more efficient than endogenously created friction in overcoming double moral hazard problems in markets.

C. Empirical Predictions and Policy Implications

Our model generates several testable empirical implications. The first implication is about the 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

15. Condition (18) depends on the endogenous variable E^*. A more primitive condition cannot be derived unless we impose a specific functional form on $p(E)$. This is because E^* cannot be explicitly solved with a general $p(E)$, which is evident from Equation (12). To derive a more primitive condition, we assume

$$p(E) = u + v/(\delta\rho) + k(L-E),$$

with $k \geq 0$ being some constant. Note that the smaller the k, the more elastic is the revenue product of labor. Now, Condition (18) can be written as:

$$v^2(1-\delta)^2/\delta^2 + (1-\rho)[v/(\delta\rho) - c](1-\delta)/\delta > kL(1-\rho)[v/(\delta\rho) - c].$$

The above condition is valid if k is small enough or the exogenous turnover rate $1-\rho$ is close to 0.

bonuses, which reduce the wage premium. This also implies that workers paid higher bonuses actually earn less than those paid low bonuses.[16]

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 level is nonmonotonic. 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. Specifically, wages are less sensitive to the employment level in occupations with high turnover costs than in occupations with low turnover costs.

By Proposition 4, if the existing turnover costs are small in markets, it might be beneficial for the government to tax employers (without compensating employees) whenever turnover occurs, thus increasing the effective turnover costs. Actually, a government tax always increases social welfare as long as the turnover costs $c < v/(\delta\rho)$. To see this, suppose $c < v/(\delta\rho)$ and the government imposes a turnover tax t with $c + t < v/(\delta\rho)$. Now, the social surplus becomes

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

The equilibrium E^* satisfies

$$p(E^*) = u + v/(\delta\rho) + [v/(\delta\rho) - c - t](1-\rho)E^*/(L-E^*).$$

Note that $\partial E^*/\partial t = \partial E^*/\partial c > 0$. Thus,

$$\partial S/\partial t = [p(E^*) - (u+v) - (1-\rho)c]\partial E^*/\partial t > 0.$$

Compared to Equation (17), the second term disappears because the government collects the money. Therefore, the total social welfare is always increasing in t as long as $c + t \leq v/(\delta\rho)$. In fact, social welfare is maximized when $t = v/(\delta\rho) - c$. Of course, the downside of a turnover tax is that it may impede the free relocation of labor when firms have different growth prospects, which is not modeled in the paper.

V. INCORPORATING SEARCH COSTS

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 jobs, regardless of whether he finds one. Although workers bear search costs directly, firms have to induce workers to search for two reasons. First, firms need workers to fill vacancies after exogenous separations occur. Second, to effectively discipline shirkers, firms have incentives to reduce the effective job acquisition rate by inducing workers to search.

We model the unemployed workers' search behavior as follows. Anticipating the wage contracts that firms offer, in each period, each unemployed worker decides whether and with what 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 make 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 σ, the stationary job acquisition rate is

$$a = \min\{1, (1-\rho)E/[\sigma(L-\rho E)]\}.$$

The presence of search costs c_S has two effects on wage contracts. First, it can discipline employed workers: if a worker shirks, he will be fired and has to incur search costs c_S for several periods to find another job. Second, firms now have to induce workers to

16. Of course, this result crucially depends on the risk neutrality of workers.

search, since search costs are directly borne by workers. Specifically, with the presence of c_S, the value functions become the following:

$$U^N = (w + b - v) + \delta[\rho U^N + (1 - \rho)\overline{U}]$$
$$U^S = w + \delta\overline{U}$$
$$\overline{U} = \max\{u + \delta\overline{U}, a \cdot U^N + (1 - a)(u + \delta\overline{U}) - c_S\}.$$

Note that $\overline{U}$ is the maximum of two payoffs: the payoff if a worker searches and the payoff if he does not search, taking a as given. 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

$$\underline{a}(E) \equiv (1 - \rho)E/(L - \rho E)$$

Specifically, a is determined by the following formulas:

$$a = \begin{cases} \underline{a} & \text{if } \underline{a}[U^N - (u + \delta\overline{U})] - c_S > 0 \quad (C1) \\ 0 & \text{if } U^N - (u + \delta\overline{U}) - c_S < 0 \quad (C2) \\ c_S/[U^N - (u + \delta\overline{U})] & \text{if } c_S \leq U^N - (u + \delta\overline{U}) \leq c_S/\underline{a} \quad (C3) \end{cases}.$$

In the first case, each unemployed worker searches with $\sigma = 1$. This is because the return of search is higher than the search costs even if all unemployed workers search with $\sigma = 1$. In the second case, each unemployed worker does not search ($\sigma = 0$) because the return of search is too low. In the third case, neither Condition (C1) nor Condition (C2) is satisfied. In this situation, only a mixed strategy equilibrium exists: each unemployed worker searches with $\sigma \in (0, 1)$ such that everyone is indifferent between searching and not searching. More explicitly, Condition (C3) and a can be written as:

$$c_S(1 - \delta\rho) \leq (w + b) - (u + v) \leq (1 - \delta\rho)c_S/\underline{a} \tag{19}$$

$$a = c_S(1 - \delta\rho)/[(w + b) - (u + v)]. \tag{20}$$

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

$$\min_{w,b}\{w + b\}$$

subject to:

$$U^N \geq \overline{U} \quad (\text{IRW1})$$
$$a[U^N - (u + \delta\overline{U})] - c_S \geq 0 \quad (\text{IRW2})$$
$$\overline{V} = V^N - c = 0 \quad (\text{IRF})$$
$$U^N \geq U^S \quad (\text{ICW})$$
$$V^N \geq V^C \quad (\text{ICF}).$$

Compared to the programming problem in the basic model, the IRW2 condition 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:

$$U^N - (u + \delta\overline{U}) \geq -c_S/(1 - a).$$

The IRW2 condition can be rewritten as:

$$U^N - (u + \delta\overline{U}) \geq c_S/a.$$

Thus, the IRW1 condition is redundant. Getting rid of the value functions, we can simplify the programming problem as follows:

$$\min_{w,b}\{w + b\}$$

subject to:

$$w \geq u + v - b + (1 - \delta\rho)c_S/a \quad (\text{IRW2})$$
$$w \geq u + (v - b - \delta\rho c_S)/[(1 - a)\delta\rho] \quad (\text{ICW})$$
$$b \leq \delta\rho c \quad (\text{ICF}).$$

Depending on parameter values, we have two possible scenarios. In the first scenario, $\delta\rho(c + c_S) \geq v$, thus efficiency wages are not necessary. Without loss of generality, firms set $b^* = v - \delta\rho c_S$. Now, the ICW condition becomes $w \geq u$, and the IRW2 condition becomes

$$w \geq u + \delta\rho c_S + (1 - \delta\rho)c_S/a.$$

Thus, the ICW condition is redundant if the IRW2 condition is satisfied. In the optimal contract, the IRW2 is binding with $a = 1$. That is,

$$w^* = u + \delta\rho c_S + (1 - \delta\rho)c_S/a = u + c_S.$$

Under the optimal contract, it can be easily seen that Equation (19) is satisfied and $a = 1$ by Equation (20). This is because under the optimal contract, unemployed workers will adjust their search behavior such that the job acquisition rate $a = 1$. The ALC curve can be written as:

$$(21)\quad \text{ALC}(E) = w^* + b^* + (1 - \delta\rho)c = u + v + (1 - \delta\rho)(c + c_S),$$

which is independent of the employment level E.

In the second scenario $\delta\rho(c + c_S) < v$, thus efficiency wages are necessary. The optimal bonus is $b^* = \delta\rho c$. Now, the IRW2 condition no longer implies the ICW condition. The following lemma specifies the optimal w^*.

LEMMA 2. *There is a cutoff* $\hat{E} \in (0, L)$ *such that*

$$(22)\quad w^* = \begin{cases} u + v - \delta\rho c + (1 - \delta\rho)c_S(L - \rho\hat{E}) \\ /[(1-\rho)\hat{E}] \ \text{ if } E \le \hat{E} \\ u + [v - \delta\rho(c + c_S)](L - \rho E) \\ /[\delta\rho(L - E)] \ \text{ if } E > \hat{E} \end{cases},$$

where $\hat{E}$ *is defined by*

$$(23)\quad v - \delta\rho c + c_S(1 - \delta\rho)(L - \rho\hat{E})/[(1-\rho)\hat{E}] = [v/(\delta\rho) - c - c_S](L - \rho\hat{E})/(L - \hat{E}).$$

Proof. See the Appendix. ■

The intuition for Lemma 2 is as follows. With the presence of search costs, wage contracts need to serve two purposes: motivate employed workers and induce unemployed workers to search. Both of them require the wage to be high relative to unemployment benefit. Moreover, both of them are affected by the employment level. When the 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 the employment level is low, inducing search is relatively more difficult. However, because unemployed workers will endogenously adjust their search behavior, the wage is set just high enough to induce the job acquisition rate $\hat{a} = (1 - \rho)\hat{E}/(L - \rho\hat{E})$, at which both the ICW and the IRW2 conditions are binding.

Following Lemma 2, the ALC curve ALC(E) can be easily derived, and the equilibrium employment level E^* is pinned down by firms' free entry condition: $p(E^*) = \text{ALC}(E^*)$.

We are interested in the wage-employment relationship. By Equation (22), when $E > \hat{E}$, the total wage payment $w^* + b^*$ is increasing in the employment level E. On the other hand, when $E < \hat{E}$, the total wage payment $w^* + b^*$ is independent of E.[17] This is an interesting result: the wage-employment relationship changes over the course of business cycle. Specifically, wages are rigid during recessions and are positively correlated with the employment level during booms. The following proposition summarizes the above result.

PROPOSITION 4. *Suppose* $\delta\rho(c + c_S) < v$. *Wages are rigid during recessions and are positively correlated with the employment level during booms. Specifically, the total wage payment is increasing in* E *when* $E > \hat{E}$ *and is independent of* E *when* $E < \hat{E}$.

If unemployed workers play pure strategies regarding search, then the total wage payment is actually decreasing in E when $E < \hat{E}$. This is because when $E < \hat{E}$, the IRW2 condition is binding. A decrease in E directly reduces the return to searching, as the job acquisition rate a decreases. To induce every unemployed worker to search (with probability 1), the total wage payment has to increase. As a result, the wage-employment relationship now assumes a U-shape. This is a more dramatic result: wages are procyclical during booms 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.

The existing theories usually 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

17. Note that without search costs, the total wage payment is always increasing in E when efficiency wages are necessary.

and unemployment, while the migration model of Harris and Tadaro (1970) predicts that there is always a positive relationship. In contrast, our model predicts that the wage-unemployment relationship might be different in different phases of business cycle.

Contractual approaches to wage determination predict that wages are history dependent. Based on the implicit contract approach of Harris and Holmstrom (1982), Beaudry and DiNardo (1991) showed that wages are downward rigid and are bid up if the market condition improves sufficiently to ensure that workers do not quit.[18] Thus, wage payments depend on the most favorable labor market condition observed since one has begun his job. MM (1993) and Malcomson (1997) studied the dynamics of fixed wage contracts in the presence of holdup problems. They show that wages are rigid with respect to shocks of small magnitudes. When the cumulative shock reaches sufficient magnitude, the wages are renegotiated either upward or downward to reflect the current market condition. Though related, our prediction is different from those implications. Specifically, in our model, wages are flexible (both upward and downward) in booms ($E > \hat{E}$) but are rigid in recessions ($E < \hat{E}$). On the other hand, the history-dependent wage-employment relationships predicted by the above papers do not depend on whether the economy is in recession or boom.

We are also interested in how changes in turnover costs affect the wage rigidity region. From Equation (23), we can see that an increase in c or c_S reduces $\hat{E}$; thus, the wage rigidity region expands. Intuitively, an increase in c or c_S reduces the efficiency wage that is necessary to motivate workers, thus inducing workers to search becomes relatively more difficult. As a result, the job acquisition rate $\hat{a}$ (hence $\hat{E}$) at which both the ICW and the IRW2 conditions are binding increases. By Equation (22), an increase in c or c_S makes wages in the wage-procyclical region less sensitive to the employment level. This is simply because the required wage premium is smaller.

18. In a model of long-term implicit contracts, Harris and Holmstrom (1982) showed that wages are downward rigid and are bid up when workers' perceived ability increases. These results crucially depend on firms' ability to commit to long-term contracts.

VI. EMPIRICAL EVIDENCE

Tables 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.

A. Forms of Contracts

There are only a few empirical works on the relationships between occupation and bonuses.[19] And in available data, only the frequency of bonuses is reported but not the amount of the bonuses. Note that the prediction of our model is about different amounts 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.

Table 3 is excerpted from table 3B of MacLeod and Parent (1999), which is 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-sized bonuses. A similar pattern holds when bonus plus promotion is used as the measure of discretionary pay.[20] This pattern is largely consistent with the predictions of our model. This is because managers usually have high turnover costs, service workers low turnover costs, and professionals and secretaries medium turnover costs.

The turnover costs for managers are usually 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 personality, so finding appropriate candidates for a vacancy takes a 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

19. Throughout this section, both merit pay and bonuses refer to subjective performance pay. They are not objective performance pay, which is conditional on contractible performance.

20. Promotion, which usually is associated with a permanent wage increase, as a deferred compensation can also motivate workers to exert effort.

TABLE 1
Empirical Predictions I

	Low Turnover Costs	High Turnover Costs
Contract form	Mainly efficiency wages	Big bonuses
Total wage payment	High	Low
Unemployment	High	Low
Wage-unemployment elasticity	High	Low

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'.

A 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 is 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.

Though this empirical evidence is largely consistent with our model, it is not a test of our theory.[21] We hope that some carefully designed empirical work can be done in the near future to directly test our model.

B. Wage Differentials and Unemployment

Using data from the Industrial Wage Survey, Brown (1992) conducted an empirical study on the relationship between wage levels and methods of pay. He found that workers paid by standard rates on average earn a higher wage than those 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.[22]

One implication of our model is that the level of equilibrium unemployment is a decreasing function of the usage of bonuses. Based on the 1990 British WIRS data, MM found 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 (1999) further supports this result. Using data from the NLSY 1988–1990, they showed that there is a strong negative relationship between the use of discretionary bonuses and local unemployment.

C. The Wage Curve

As in Solon, Barsky, and Parker (1994), most of the empirical works just test whether the 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 U.S. General Social Survey (GSS) data (1974–1988), they estimated the industry wage curve (wage as a function of the unemployment rate in industries) and the regional wage curve (wage as a function of the regional unemployment rate). Figure 3, copied from BO (p. 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

21. One may think that the pattern given in Table 3 is also consistent with firms using bonuses to select more able workers, which may matter more in managerial occupations and for professionals. However, bonuses cannot be enforced by the court. Therefore, how much bonus will be posted does not only depend on how much bonus firms are willing to post but also depend on how much firms can credibly post.

22. Combined with the information given in Table 3, the prediction that occupations receiving higher bonuses would have a lower level of total compensation implies that managers would have lower total compensation on average than cleaning service workers. This seems to contradict the fact that managers earn more than cleaning service workers. However, in Brown's empirical analysis, he controlled for human capital. The fact that managers earn more than cleaning workers is simply because the former have more human capital. After controlling for human capital, cleaning workers actually earn more than managers.

TABLE 2
Empirical Predictions II

The wage curve	Procyclical in booms, rigid in recessions	
	Low Turnover Costs	High Turnover Costs
The wage rigidity region	Big	Small
Wage-unemployment elasticity	High	Low

of the wage curve suggests that unemployed workers are not able to adjust perfectly their search behavior.

BO also estimated the wage curve based on the U.S. Current Population Surveys (1964–1991). The results are slightly different from those from the GSS: the wage curve is significantly downward sloping when the unemployment rate is low, and it flattens out as the unemployment rate increases, but there is no upward-sloping portion of the wage curve. A similar estimation result holds for data from the British Social Attitude Surveys (1983–1989): the wage curve flattens out when the unemployment rate is greater than 13%. Using the General Household Surveys' data (1973–1977) from Britain, BO found that the wage curve has a U-shape with the turning point occurring around an unemployment rate of 4.5%. Based on the International Social Survey Program data (1986–1991) from West Germany, the estimation of BO shows that the wage curve flattens out around an unemployment rate of 11%. Bratsberg and Turunen (1996) estimated the U.S. wage curve for young workers using the 1979–1993 waves of the NLSY. According to their study, the wage curve based on annual earnings flattens out when the unemployment rate is higher than 12% and the wage curve based on hourly wage exhibits a U-shape with the minimum wage reached at an 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 rates. They are largely consistent with our empirical prediction that wages are procyclical during booms and either rigid or countercyclical in recessions.

D. Wage-Unemployment Elasticity

BO also estimated the wage-unemployment relationship for different occupations using data from the 1990 British WIRS. The results are reported in Table 5. Unskilled manual workers have the largest wage-unemployment elasticity, supervisors the lowest elasticity, and clerical workers 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 leads to low wage-unemployment elasticity.

Table 6 reports the results of two other studies: BO on the British General Household Surveys 1973–1977 and Kennedy and Borland (2000) on the Australian Bureau of Statistics Income Distribution Survey 1982–1994. Again, managers have the lowest wage-unemployment elasticity, and manual workers or clerks have the highest elasticity.

TABLE 3
Bonuses across Occupations I

Occupations	Bonus (%)	Bonus + Promotion (%)
Managers	28.46	47.37
Professionals	15.46	29.22
Secretaries	11.60	25.60
Food service workers	7.49	18.66
Cleaning service workers	7.43	17.33

TABLE 4
Bonuses across Occupations II

Occupations	Incidence of Bonus (%)
Professional and managerial	35
Supervisors	32
Clerical, administrative, and secretarial	30
Skilled manual	22
Semiskilled manual	16
Unskilled manual	11

FIGURE 3
The Wage Curves

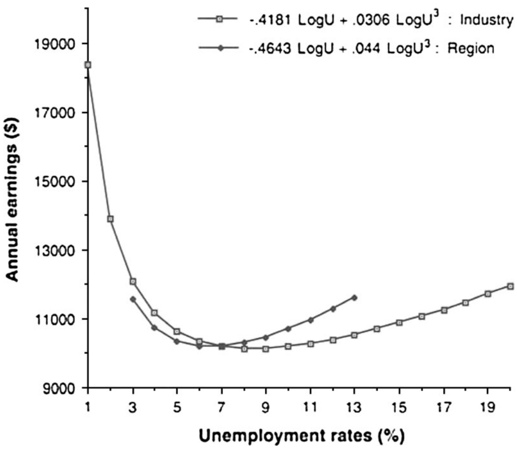


TABLE 6
Wage-unemployment Elasticities II

Occupations	British GHS data	ABS IDS data
Manual	−0.0721	
Clerks		−0.0896
Professionals	−0.0631	−0.0224
Managers	−0.0497	−0.0198

Notes: GHS = General Household Survey; ABS IDS = Australian Bureau of Statistics Income Distribution Survey.

VII. 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 the firms' incentive to renege. The presence of turnover costs borne by firms reduces firms' incentives to renege, thus making implicit bonuses credible to some extent. In the optimal contracts, the amount of the bonus is positively correlated and the amount of wage premium 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 increasing the equilibrium employment level and social welfare.

The extended model incorporates workers' search costs. In this setting, the 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 occupations with different 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 consistent with these predictions.

TABLE 5
Wage-unemployment Elasticities I

Occupations	Coefficient
Unskilled manual	−0.0916
Skilled manual	−0.0325
Clerical	−0.0434
Supervisors	−0.0048

Though our model is couched in a labor market setting, it can also be applied to other markets where both parties in a relationship have moral hazard problems and both are able to change partners in markets. For example, consider buyer-seller relationships in a market setting, in which the quality of goods is observable but not verifiable. To motivated sellers, buyers can either offer higher fixed prices (analogous to efficiency wages) or post some bonuses (which have to be self-enforcing), tying payments to the quality of goods. Turnover costs will generally affect the optimal contracts and have welfare implications, similar to those shown in the labor market model.

The model can be extended in several directions. First, in the model, we have assumed that the turnover costs are exogenous. When firms are able to choose turnover costs within some range, they might have incentives to choose the level of turnover costs that minimizes their ALCs. Second, in the model, we have assumed that workers are homogeneous. Yang (2005) studied relational contracts with heterogeneous workers. Now moral hazard interacts with firms' learning about workers' types, which results in nonstationary relational contracts. This helps to explain why contractual terms change as the length of a relationship increases. Third, it is also interesting to model relational contracts in a setting where the demand for labor is fluctuating over time. Such a model enables us to see more explicitly how wages and contract forms change during the course of business cycles. This is left for future research.

APPENDIX

Proof of lemma 1.

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

$$\min_{w\geq 0,\ b\geq 0} \{w+b\}$$

$$(24)\qquad w+b\geq u+v \quad \text{(IRW)}$$

$$(25)\qquad w\geq u+k(v-b) \quad \text{(ICW)}$$

$$(26)\qquad b\leq \delta\rho c \quad \text{(ICF)}$$

where $k = 1/[(1-a)\delta\rho] > 1$. Condition (25) can be further reformulated as:

$$(27)\qquad w+b\geq u+v+(k-1)(v-b).$$

From Equation (27), we see that if $b < v$, then only Equation (25) is binding and the total compensation $w + b$ is decreasing in b; so b should be set as high as possible subject to condition (26). Condition (24) is binding if and only if $b \geq v$. Actually, Condition (24) specifies the low bound of the total wage payment, $u + v$. Therefore, when $c < v/(\delta\rho)$, the optimal contract is $b^* = \delta\rho c$ and $w^* = u + k(v - b^*)$. When $c \geq v/(\delta\rho)$, the optimal contracts are $b^* \in [v, \delta\rho c]$ subject to $w^* = u + v - b^* > 0$. ■

Proof of lemma 2.

Note that the right-hand side of the IRW2 condition is decreasing in a and goes to infinity as a goes to 0. On the other hand, the right-hand side of the ICW condition is increasing in a and goes to infinity as a goes to 1. Thus, there is an $\hat{a} \in (0, 1)$ such that the right-hand sides of both conditions are equal. That is,

$$(28)\qquad v-\delta\rho c+(1-\delta\rho)c_S/\hat{a} = (v-\delta\rho c-\delta\rho c_S)/[(1-\hat{a})\delta\rho]$$

Therefore, if $a > \hat{a}$, then the IRW2 condition is redundant and the ICW condition is binding. If $a \leq \hat{a}$, then the ICW condition is redundant and the IRW2 condition is binding. Moreover, the fixed wage w is minimized when $a = \hat{a}$. Define $\hat{E}$ such that

$$(29)\qquad (1-\rho)\hat{E}/(L-\rho\hat{E}) \equiv \hat{a}.$$

Note that $\hat{E}$ is unique and $\hat{E} \in (0, L)$. Combining Equations (28) and (29) gives rise to Equation (23).
First, consider the case $E > \hat{E}$. Then,

$$\underline{a}(E) = (1-\rho)E/(L-\rho E) > (1-\rho)\hat{E}/(L-\rho\hat{E}) = \hat{a}.$$

Thus, the binding condition is the ICW condition, and firms will induce unemployed workers to search with probability 1 with the effective a equaling the low bound $\underline{a}(E)$. As a result, the optimal w^* is given by:

$$w^* = u+[v-\delta\rho(c+c_S)](L-\rho E)/[\delta\rho(L-E)].$$

Now, consider the case $E \leq \hat{E}$. In this case, firms will just induce $\hat{a}$ by paying the following wage:

$$w^* = u+v-\delta\rho c+(1-\delta\rho)c_S(L-\rho\hat{E})/[(1-\rho)\hat{E}].$$

With the wage specified above, it can be verified that Equation (19) is satisfied and that from Equation (20), the effective a is

$$a = c_S(1-\delta\rho)/[(w^*+\delta\rho c)-(u+v)] = (1-\rho)\hat{E}/(L-\rho\hat{E}) = \hat{a}.$$

REFERENCES

Akerlof, G., and L. Katz. "Workers' Trust Funds and The Logic of Wage Profiles." *Quarterly Journal of Economics*, 104, 1989, 525–36.

Baker, G., R. Gibbons, and K. Murphy. "Subjective Performance Measures in Optimal Incentive Contracts." *Quarterly Journal of Economics*, 109, 1994, 1125–56.

Beaudry, P., and J. DiNardo. "The Effects of Implicit Contracts on the Movement of Wages over the Business Cycle: Evidence from Micro Data." *Journal of Political Economy*, 99, 1991, 665–88.

Blanchflower, D., and A. Oswald. The Wage Curve. Boston, MA: The MIT Press, 1994.

Bratsberg, B., and J. Turunen. "Wage Curve Evidence from Panel Data." *Economics Letters*, 51, 1996, 345–53.

Brown, C. "Wage Levels and Method of Pay." *Rand Journal of Economics*, 23, 1992, 366–75.

Bull, C. "The Existence of Self-Enforcing Implicit Contracts." *Quarterly Journal of Economics*, 102, 1987, 147–59.

Farrell, J., and E. Maskin. "Renegotiation in Repeated Games." *Games and Economic Behavior*, 1, 1989, 327–60.

Green, F., A. Felstead, K. Mayhew, and A. Pack. "The Impact of Training on Labor Mobility: Individual and Firm-Level Evidence from Britain." *British Journal of Industrial Relations*, 38, 2000, 261–75.

Harris, M., and B. Holmstrom. "A Theory of Wage Dynamics." *Review of Economic Studies*, 49, 1982, 315–33.

Harris, J., and M. Tadaro. "Migration, Unemployment and Development: A Two-Sector Analysis." *American Economic Review*, 60, 1970, 126–42.

Hashimoto, M., and B. Yu. "Specific Capital, Employment Contracts, and Wage Rigidity." *Bell Journal of Economics*, 11, 1980, 536–49.

Kennedy, S., and J. Borland. "A Wage Curve for Australia." *Oxford Economic Papers*, 52, 2000, 774–803.

Lang, K. "Persistent Wage Dispersion and Involuntary Unemployment." *Quarterly Journal of Economics*, 106, 1991, 181–202.

Levin, J. "Relational Incentive Contracts." *American Economic Review*, 93, 2003, 835–57.

MacLeod, B., and J. Malcomson. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." *Econometrica*, 57, 1989, 447–80.

———. "Investments, Holdup, and the Form of Market Contracts." *American Economic Review*, 83, 1993, 811–37.

———. "Motivation and Markets." *American Economic Review*, 88, 1998, 388–411.

MacLeod, B., and D. Parent. "Job Characteristics and the Form of Compensation." *Research in Labor Economics*, 18, 1999, 177–242.

Malcomson, J. "Contracts, Holdup, and the Labor Markets." *Journal of Economic Literature*, 35, 1997, 1916–57.

Salop, S. "A Model of the Natural Rate of Unemployment." *American Economic Review*, 69, 1979, 117–25.

Shapiro, C., and J. Stiglitz. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, 74, 1984, 433–44.

Solon, G., R. Barsky, and J. Parker. "Measuring the Cyclicality of Real Wages: How Important is Composition Bias." *Quarterly Journal of Economics*, 109, 1994, 1–25.

Stiglitz, J. "Alternative Theories of Wage Determination and Unemployment in LDCs: The Labor Turnover Model." *Quarterly Journal of Economics*, 88, 1974, 194–27.

Yang, H. "Nonstationary Relational Contracts." Working Paper, Ohio State University, 2005.