NBER WORKING PAPER SERIES

DO DEFERRED WAGES DOMINATE INVOLUNTARY UNEMPLOYMENT AS A WORKER DISCIPLINE DEVICE?

George A. Akerlof

Lawrence F. Katz

Working Paper No. 2025

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
September 1986

We thank Katharine Abraham, William Dickens, Kevin Lang, Kevin Murphy, Lawrence Summers, Andrew Weiss, Janet Yellen and participants in seminars at the NBER Summer Institute for helpful comments. Phillip Bokovoy provided expert research assistance. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the authors and not those of the National Bureau of Economic Research.

NBER Working Paper #2025
September 1986

Do Deferred Wages Dominate Involuntary Unemployment as a Worker Discipline Device?

ABSTRACT

In the most widely analyzed type of efficiency wage model of involuntary unemployment, firms pay wages in excess of market clearing to give workers an incentive not to shirk. Such payments in excess of market clearing and the resultant equilibrium unemployment act as a worker discipline device. This paper concerns what is usually considered the most important theoretical criticism of such models: the so-called bonding argument. The essence of the bonding critique is that contracts whereby workers pay a bond to the firm upon taking a job (or pay an employment fee to gain employment) can eliminate involuntary unemployment. Explicit upfront bonds are only quite rarely observed. A more subtle form of the bonding critique argues that implicit bonding through upward sloping wage profiles and other deferred payment schemes can perfectly substitute for upfront bonds in providing incentives not to shirk and thereby allow the labor market to clear.

This paper shows that upward sloping wage profiles do not act as a perfect substitute for explicit bonds in a natural extension of the shirking model in which workers are finite lived, the monitoring of worker behaviors on the job is costly, and firms have reputations for honesty as employers. In the absence of direct upfront bonding, optimal payment schedules will be in excess of market clearing. The reason why upward sloping wage profiles that are market clearing will not generally be the optimal labor contract is simple: delayed payment may provide sufficient incentive to prevent shirking late in the life of the contract, but in the beginning of the contract it does not prevent shirking. And it turns out in a variety of stylized cases, it is cheaper for the firm to pay a wage premium rather than to accept worker shirking early in the contract. The implications of potential worker malfeasance in the absence of explicit bonds for compensation schedules, job assignments, and firm monitoring strategies over the course of a worker's career are also analyzed.

George A. Akerlof
Department of Economics
Univesity of California
Berkeley, Ca. 94720

Lawrence F. Katz
Department of Ecomomics
Harvard Univeristy
Cambridge, Ma. 02138

I. INTRODUCTION

In the most popular type of efficiency wage model of involuntary unemployment, firms pay wages in excess of market clearing to give workers an incentive not to shirk.[1] Such payments in excess of market clearing and the resultant equilibrium unemployment act as a worker discipline device. This paper concerns what is usually considered the most important theoretical criticism of such models: the so-called bonding argument.

Most efficiency wage models of the worker-discipline type explicitly assume contracts where remuneration consists only of a constant wage. The wage is unable to clear the labor market when it must simultaneously act both to allocate labor and to provide incentives for adequate employee performance. However, more complicated contracts whereby workers pay a bond to the firm upon taking a job (or pay an employment fee to gain employment) could eliminate involuntary unemployment.[2] The threat of forfeiting the bond or paying a new employment fee to acquire a job can generate work incentives allowing the wage to adjust so that the total terms of the labor contract clear the labor market.

Below we shall present a model in which workers have a choice whether to work or shirk, firms have reputations for honesty, and there are no net payments by workers to firms. We shall explain in turn why each of these assumptions is natural in view of empirical realities and recent work on efficiency wage theory and the bonding problem.

1. Work-shirk: At least since the movement away from Taylorism, human resource management and industrial relations specialists have emphasized the extent to which workers are imperfectly controlled and monitored by their employers. We note two recent exemplary studies which illustrate this point. The remarkable study Cheats At Work by Mars (1982) shows the extent to which workers manage to cheat, suggesting the importance of workers' discretion over their activities. Mars documents a wide variety of workplace "fiddles" ranging from the practice of supermarket checkout clerks of ringing up less than the proper amount and pocketing the difference to the techniques developed by bus drivers to speed up their journeys and gain longer breaks by picking up as few passengers as possible. One dramatic but typical example concerns longshoremen in the port of St. John's, Newfoundland. The dock gangs had systematic techniques for generating "accidental" breakage in cargo so that it could be "legitimately" diverted permitting the stowers to pilfer the cargo shielded from the view of supervisors. Such an incident is described in detail as follows:

> The pilferage of thirteen of a consignment of a hundred men's suits provides a good example of this unofficial history of events. The hatch-checker knew from his bills of landing what sort of cargo was in transit. The vessel team then engineered an 'accidental' fall of two crates. The damaged crates were moved normally with the rest of the cargo to the shed. The shed team then engineered the undisturbed pilferage of the suits. To do this the fork-lift truck driver had to stack other cargo in such a way as to block off what was happening from the superintendent's line of vision. At the same time, he constructed by skillful stacking, a changing-room where suits could be tried on by the men. The next day the hatch-checker reported that two crates were missing. By then the suits had disappeared -- smuggled out of the dock under the men's clothes. (Mars 1982; pp. 104-5).

Roy's (1952) study and later Burawoy's (1979) (coincidentally at the same piecework machine shop) document in detail the ways in which workers via their own individual initiatives and their social interactions manage to obtain autonomy relative to management. The following example captures the flavor of

worker-management relations in this plant. Burawoy (citing the Roy study) reports an attempt by management in 1945-46 to prevent workers from signing onto a new job before signing off their previous one. The workers were no longer allowed to collect prints and fixtures for their new job before presentation of their yellow cards signing off their old jobs. The initial response by workers to this changed procedure was to steal into the crib for prints and fixtures. Management's response of preventing all workers but the crib attendant to enter the crib slowed output greatly and in consequence the new regulations were quickly rescinded.

The observation that workers are often only imperfectly monitored is incorporated into our model by assuming that workers have a choice whether to work or shirk and that the probability of being caught shirking is less than one. This assumption is of course not new: it is at the heart of the unemployment as a worker discipline device model.[3]

2. Firm reputation and honesty: If firms as well as workers are not honest, the bonding argument may not be a valid reason why labor markets clear: if firms are not trustworthy workers would be hesitant to post a bond and risk the chance that dishonest firms will dismiss them and retain the bond. Stoft (1982) and more recently MacLeod and Malcomson (1986) show that if firms are not trustworthy, wages may be greater than market clearing and there will be equilibrium involuntary unemployment.[4] It is surely true that the opportunistic behavior modeled by these authors is empirically observed and important; but we also believe that a great deal of business is repeat business in which reputations can be easily established and there is a return to honesty. If all firms were untrustworthy and as a consequence had to pay more than the market clearing wage to attract workers, an entrant with a reputation for honest treatment of workers would be able to attract

labor at their reservation price. Contracts would be of the form where workers would initially post a bond to the firm and then collect a wage in excess of their reservation wage. The value of the remuneration package would be just the reservation price of the marginal worker. In this way because the honest firm pays lower total remuneration (net of the bond posted by the worker)[5], it makes higher profits than the dishonest firms: there is a positive return to reputation. Thus there is a theoretical argument for considering labor contracts between honest firms and potentially shirking workers.

Our major concern with resting the case for efficiency wages on a model assuming honest firms is empirical. It is our feeling that in most repeat business firms do pay considerable attention to their reputation. Probably the most famous study of this was made by Macaulay (1963) whose study of business practices by 43 companies in Wisconsin showed that these firms were in many respects willing to give customers terms which were better than guaranteed by their legal contracts. Orders could be canceled; adjustments were made for defective merchandise without question, etc. In a more extraordinary example which is exactly germane to the labor market Allen, Clark, and Sumner (1984) present evidence that many firms with no legal obligation to do so have voluntarily paid <u>retirees</u> inflation adjustments above those to which they were legally entitled. Such payments are almost certainly explained by firms' concerns with their reputations as fair employers. Allen, Clark, and Sumner find that 75 percent of all beneficiaries in their sample received at least one post-retirement increase in pension benefits between 1973 and 1979.

The recent work by Axelrod (1984) suggests a quasi-empirical reason why honesty may dominate markets where there are repeat dealings and reputation may be im-

portant. Axelrod invited contestants to write programs to play a repeated prisoner's dilemma game in a computer tournament. All entries played against each other. "Nice" programs did overwhelmingly best, where "nice" programs are those which are never the first to defect from a cooperative strategy. The clear-cut winner of the tournament was the TIT FOR TAT strategy which starts with a cooperative strategy, and thereafter does what the other player did on the previous move. The dramatic success of TIT FOR TAT in Axelrod's tournament suggests that it pays firms to have reputations for honest behavior. The TIT FOR TAT program won both rounds of Axelrod's tournament in large part because, of all the nice programs, its behavior was the least ambiguous, and thus it most easily gained a reputation for cooperative (but also retaliatory) behavior (Axelrod 1984, 20-21). Axelrod's example of common implicitly understood mutual truces between soldiers involved in trench warfare in World War I gives convincing empirical evidence of the development of reputation and cooperation in circumstances where communication is difficult.

3. Absence of Explicit Bonding: The third property of our model concerns payments by workers to firms. The model assumes that there are no net payments in any period from workers to firms. Of course this contradicts the suggestion that optimal contracts involve workers posting of bonds with the whole bond-cum wage profile being at market clearing. But empirically net payments by workers to firms to secure jobs seems very rare -- in fact sufficiently rare that the $250 to $1,000 recently paid upfront by flight attendant trainees to TWA for a training course is newsworthy (San Francisco Chronicle March 17, 1986; p. 19). Because such payments by workers to firms are so very rare the model of this paper will not allow net payments by workers to firms. What we wish to see, as will be explained

presently, is whether in the absence of explicit bonding whereby workers make payments to firms more subtle forms of bonding will not take place.

It has been suggested that bonding need not occur by the explicit posting of a bond at the start of a job, but instead by initial wage payments which are less than workers' productivity and later wage payments in excess of workers' productivity. According to this folk argument, if this is the form of the payment profile, workers in their early stages of their careers will not shirk because, if caught, they must forfeit the high earnings of later years which are in excess of their marginal product (and of their opportunity costs). It has been believed that upward sloping wage profiles (such as described by Lazear (1979, 1981) to explain mandatory retirement and hours constraints) will provide workers with an incentive structure which will prevent shirking even in the absence of a wage profile whose overall present discounted value is greater than market clearing.[6]

This paper shows that the preceding conjecture is usually false in the most natural framework of the shirking model. In the absence of upfront bonding (or explicit employment fees), optimal payment schedules will be in excess of market clearing.[7] Thus the empirical absence of payments by workers to firms cannot be explained by the substitution of implicit bonding (via upward sloping wage profiles) for explicit bonding. The preceding sentence has been italicized (underscored) for emphasis because it represents the key point of this paper. We show below with some generality that in the absence of explicit upfront bonding optimal remuneration packages will have a value to workers in excess of their opportunity costs. The presumption cannot be made that real world contracts are fully bonded unless explicit upfront bonds are directly observed. On the contrary, as our examples seem to document, in the absence of upfront bonds, in situations where

shirking and imperfect monitoring are important, the presumption should be made that compensation packages are in excess of market clearing.

II. EXPLANATION OF THE MODEL AND PLAN OF THE PAPER

The reason why upward sloping wage profiles that are market clearing will not generally be the optimal labor contract is simple: delayed payment may provide sufficient incentive to prevent shirking late in the life of the contract, but in the beginning of the contract it does not prevent shirking. And it turns out in a variety of stylized cases, it is cheaper for the firm to pay a wage premium rather than to accept worker shirking early in the contract.

A. The Trust Fund Concept

Why is there shirking at the beginning of a market clearing contract? View the firm as setting aside a trust fund of unpaid wages for the individual worker in the case of an upward sloping (but market clearing) wage profile. The worker who decides to shirk is risking giving up the trust fund. In a continuous time setting, suppose that a worker who shirks for a short time dt will be caught and lose his trust fund with probability p dt. The expected loss from shirking for a risk neutral worker at time T

$$p\ dt \int_0^T x(t)\ e^{r(T-t)}\ dt$$

where x(t) is the unpaid portion of wages and r is the interest rate. The potential gain from shirking during this interval is v dt. Thus, it takes some time

before the trust fund from unpaid wages (and accrued interest) is sufficient to prevent shirking.

We have thus seen the flaw in the basic folk theorem that upward sloping wage profiles in the absence of explicit upfront bonding, will act as an incentive against shirking, with market clearing wage profiles. This argument explains why workers may not shirk late in their careers; it fails to explain why workers will not shirk early in their careers, as will in fact occur with market clearing wages.[8] It remains, however, to show that paying a premium above market clearing wages will be a cheaper way to hire effective labor units than paying market clearing wage levels with workers shirking early in their careers. Such a proposition is not true in general. But we will show that with a rather wide variety of productivity patterns, the cost of shirking by workers early in their careers with market clearing wage schedules will be greater than the cost of paying wage premiums in excess of market clearing which prevent workers from shirking entirely.

B. Plan of Paper

The paper proceeds in several easy steps. The next section will present a simple natural model to analyze the basic problem with the folk argument that deferred payment schemes allow the labor market to clear in the presence of a worker moral hazard problem. Firms have reputations for honesty and do not cheat on their potential wage promises. Workers have an opportunity to shirk. If workers shirk, they have a probability p dt of being caught in the short interval dt, during which time their shirking will yield them utility in the amount v dt. Workers' productivity is constant at e^* over the period if working, and is 0 if shirking. Their wage working in the secondary sector (in which there is no shirking problem) is

w^* (independent of their age). The model up to this point corresponds exactly to a continuous time analogue of the Becker and Stigler (1974) model except that we add one more constraint: by assumption there is no upfront bond - presumably because workers face a capital market constraint and cannot borrow to post such a bond. In section III, the model is presented with constant w^*, constant e^*, and no discounting. The essence of the argument is highlighted in this stripped down version of the model.

In section IV, we deal with several complications to the model: will the optimal contract still be nonmarket clearing if there is discounting? if $w^*(t)$ varies over time? if $e^*(t)$ varies? The answer is not always yes. It turns out that if $e^*(t)$ rises (as seems natural) relative to $w^*(t)$, then the market clearing contract may sometimes be optimal.

At first glance a rising $e^*(t)$ relative to $w^*(t)$ may seem realistic. Workers in the primary sector are likely to acquire firm specific knowledge and human capital with work experience. Yet, workers may not simply gain specific human capital automatically through being on the job regardless of their work efforts. An alternative model in this spirit is suggested: $e^*(t)$ does not rise exogenously, it only rises if the worker is not shirking. In other words, $e^*(t)$ is a function of time actually worked, not just time on the job. In this case, it turns out that nonmarket clearing contracts are optimal. Such models are presented in section V.

Sections III, IV, and V all deal with models in which workers and firms have the same discount rate. Although it is possible in such models to show that optimal labor contracts are not market clearing, there are still many wage profiles of

equivalent present discounted value for which the firm and worker are indifferent. Introduction of a bit of myopia on the part of workers in terms of a slightly higher discount rate gives a determinate solution to the path of remuneration. Models with worker myopia are analyzed in section VI.

The systematic life cycle pattern of the incentives to shirk under deferred payment schemes calls into question the assumption that workers will be monitored in such a way that the probability of being caught shirking is constant over the course of their careers. The implications of allowing monitoring decisions to be endogenous are analyzed in section VII. The optimal monitoring strategy involves intensive monitoring early in a worker's career and less intense monitoring as the worker gains experience with the firm and builds up a larger trust fund. The optimal contract still involves an efficiency wage premium when monitoring is allowed to vary under a monitoring technology in which the likelihood of detecting worker malfeasance is proportional to monitoring expenditures.

III. BASIC MODEL

This section constructs a natural simple model to illustrate why labor markets fail to clear in the absence of employment fees even in markets where firms are trustworthy. In this model, workers have discretion over their own effort and firms have imperfect abilities to monitor shirking. After a firm observes a worker shirking, any consequent dismissal of the worker for that violation is assumed to be immediate.[9] In this environment, different compensation profiles

have different costs to the firm and also different incentives to workers. The properties of compensation profiles with the best trade off between cost and work incentives (from the firm's point of view) will be examined.

A. The Model's Assumptions

The following assumptions fully describe the model. Asterisks denote where modifications will be later discussed.

Time

1. A worker has a work career beginning at time 0 and ending at time n.

2. Time is continuous in the model.

The Work-Shirk Decision and its Consequences

3. At each point of his career, the worker has a decision whether to work or to shirk. The worker makes this decision at each point of time to maximize expected lifetime utility.

4. A worker who shirks will supply 0 units of effective labor to the firm. A worker who works will supply e* units of effective labor to the firm.[10]

5.* e* is constant throughout the worker's career.

6. The monetary value of shirking to the worker for the short length of time dt is v dt.

The Monitoring Process and Worker Discipline/ Firm Honesty in Dismissal

7.* A worker who shirks for the period dt is detected by the firm with probability p dt.

8. A delay cannot occur between observation of shirking and a worker's consequent dismissal.

9. Firms are totally honest in their dismissals. Workers are never dismissed unless caught shirking.

Alternative Opportunities for Workers

10. A worker has outside opportunities which pay a wage w*(t), for $0 \leq t \leq n$. It is convenient to think of these alternative opportunities as the secondary labor market.

11.* w*(t)is constant throughout the worker's career at w*.

12. Upon leaving the firm at t, the worker can immediately earn w*(t).

Workers' and Firms' Utilities and Discount Rates

13. Both firms and workers are risk neutral.

14.* Both workers and firms have a zero rate of discount.

15. Workers are homogeneous.

Restriction on Compensation Schedules

16. Workers do not pay firms an explicit bond upon joining the firm. In other words, there are no net payments by workers to firms.

B. Brief Comparison of the Model with the Shapiro-Stiglitz Model of Unemployment As a Worker Discipline Device and Lazear's Model of Mandatory Retirement

The major innovation in our model over that of Shapiro and Stiglitz (1984) occurs in the explicit modelling of the workers career path of wages within the firm. Shapiro and Stiglitz assume infinitely lived workers and implicitly assume the compensation path, $w(t)$, is constant. Our model may appear different from that of Shapiro and Stiglitz because it is assumed that the outside wage $w^*(t)$ is instantaneously available to a worker who is caught shirking and consequently dismissed. Such an appearance of difference, however, is superficial. In the context of Shapiro and Stiglitz, the present discounted value of $w^*(t)$ between t and n can be interpreted as the monetary value of the expected discounted stream of earnings and leisure of a worker between the date of his dismissal and his retirement. It has been widely conjectured that explicit modelling of workers' career compensation paths will show that labor markets clear in the Shapiro-Stiglitz model because upward sloping wage profiles provide incentives not to shirk since dismissal results in loss of future earnings in excess of workers' alternative opportunities.

Finally, we wish to note that our model deals explicitly with the issue analyzed by Lazear (1979, 1981) of the design of an optimal wage profile as an agency

problem. There are many differences between Lazear's model and our own, but none of them are fundamental. In Lazear's model, an employee has a chance to take an amount θ at any time t. That person will be caught with certainty in the next instant. Lazear then asks what the market wage contract will be given potential moral hazard on both sides of the market, on the side of the worker who may shirk and of the firm which may fail to make its payments. There are three differences between Lazear's model and our own. The malfeasance of the worker has a payoff which is not a lump sum θ but is proportional to the length of time the worker shirks. The probability of a worker being caught for committing malfeasance is p dt, not equal to 1 in the short period of time dt. And, it is assumed that the firm will not cheat on its implicit contract to the worker. Apparently, none of these differences between Lazear's model and our own is fundamental, at least in one important respect: both models predict a wage package with a present discounted value that exceeds the present discounted value of the alternative wage offer $w^*(t)$ in the absence of explicit employment fees paid at the time the workers begin their jobs.

C. Derivation of the Optimal Wage Path

The firm in this model wishes to purchase labor efficiency units at minimum unit cost. We shall show that the cost minimizing wage package involves total payments whose sum is $w^*n + v/p$. The alternative opportunities (which are freely available to a worker) pay a lifetime total of w^*n. Thus, total remuneration from the cost minimizing package is in excess of the total remuneration in the secondary labor market by v/p.

It is intuitive that the firm will lose nothing by paying all of the worker's remuneration at the worker's retirement date. This way the firm's expenditure on worker remuneration will do the most work in inducing workers not to shirk. At each point in his career, the worker has the inducement not to shirk of the payment at the end of his career which is only received if he is never caught shirking.

Given that we need only consider compensation schemes in which all payments are made at the end of the worker's career, it is only necessary to discover the optimal total remuneration over the worker's lifetime. The worker must be paid at least w^*n at the end of his career in order to be induced to join the firm. Suppose that the worker is paid $w^*n + x$. What is the optimum value of the premium paid to the worker above the market-clearing wage stream whose lifetime value is nw^*? In other words, what is the optimal choice of x?

Given that the firm is paying the worker w^*n+x at the end, we can view the worker's choice problem in the following way. Suppose a worker has not previously been caught shirking at time t. The worker faces the following choice at t. He may choose to shirk over the interval t to t+dt. This policy has the gain v dt due to the added utility from shirking. However, if the worker gets caught shirking, his total compensation will be $w^*(n-t)$ from future earnings in the secondary sector rather than the w^*n+x available at his firm for someone never caught shirking. Consequently, if the worker plans to work from time t+dt to n, his potential gain from shirking is v dt, and his potential loss is

$$p\ dt((w^*n+x) - w^*(n-t))$$

which simplifies to $p\ dt(w^*t+x)$.[11] At the point T where the worker is just at the margin between working and shirking, we have

(1) $$p\ dt(w^*T+x) = v\ dt.$$

At later times, it will be more costly for the worker to be caught shirking, and therefore the worker will work. And at earlier times it is less costly to be caught shirking and therefore the worker will shirk.

As described above, equation (1) suggests the simple analogy of the trust fund which underlies much of the logic of our argument. We can pretend that the firm sets up a trust fund for its workers. It puts up x in the beginning when the worker is initially hired and later puts money into the trust fund at rate w^*. At each point in time t, the worker must decide whether to shirk, with the ill consequence that he may be caught with probability p dt and give up the accumulated trust fund of amount w^*t+x. The potential gain from shirking is v dt. Consequently the worker is just indifferent between working and shirking at time T_x for which

(2) $$p\ dt\ [w^*T_x + x] = v\ dt.$$

What is the optimal value of x given that T_x obeys (2)? Equation (2) yields the value of T_x for each x,

(3) $$T_x = \min((v/p - x)/w^*, 0).$$

The firm's problem is to choose x to minimize unit labor costs which are given by

(4) $$\frac{w^*n + x}{(n-T_x)e^*} = \frac{w^*n + x}{(n-[(\frac{v}{p} - x)/w^*])e^*}$$

over the range $0 \leq x \leq v/p$. It is easily shown that expression (4) is minimized if $x=v/p$, over the range $0 \leq x \leq v/p$, since the derivative of (4) with respect to x is negative for all x in the range $0 \leq x \leq v/p$.

As a result, the optimal (cost minimizing) wage package will pay a premium $x=v/p$. This implies $T_x=0$. There is never any shirking under the optimal compensation profile and the firm makes total career payments of v/p in excess of the market.

An explanation for this solution proceeds as follows. For a worker ever to work, at the last instant worked he or she must receive a surplus of at least v/p. This v/p constitutes a fixed cost to the firm. At all previous moments worked, the worker must also have a surplus of at least v/p, so that the firm pays a minimum to the worker of w^*t_w+v/p for working a length of time t_w. By paying w^*n+v/p at the end of the worker's career, the firms spreads the fixed cost v/p over the maximum working time (the worker's whole career n) and therefore unit labor cost is minimized.[12]

It has been seen that in the case without discounting that a firm which minimizes its unit labor costs in the presence of a shirking problem and without the ability to collect upfront employment fees from workers must pay a career wage that exceeds the alternative career earnings available to workers in the secondary sector by v/p. Upward sloping wage profiles cannot fully substitute for explicit employment fees in such a model. Only after the trust fund has an accumulated value v/p will the worker stop shirking. And if "bonding" occurs by workers' initial receipt of wages below the secondary sector level, it takes too long for the worker to stop shirking if total lifetime wages paid out are at the market clearing level. It is better instead for the firm to pay an efficiency wage

premium of v/p in excess of market clearing at the end of the worker's career and prevent shirking altogether. Upward sloping wage profiles alone do not eliminate the need for efficiency wage payments when worker behavior can only be imperfectly monitored.

Remark:

The model above has only one type of job for workers in the primary sector. If jobs differ according to the ease of monitoring, a firm's optimal strategy is to assign younger workers (recent hires) to more easily monitored jobs. Indeed, if there are enough productive jobs with costless monitoring, the equilibrium contract will be market clearing with workers placed in jobs with no shirking potential early in their careers and moved to more responsible jobs once their trust funds have built up sufficiently. Furthermore, if primary sector firms are paying above market clearing wages because of monitoring difficulties, there is an incentive for primary sector firms to merge with secondary sector firms. In this case, workers would be positioned in the secondary sector jobs early in their careers and optimal deferred payment schemes could potentially be market clearing. The importance of job specific human capital accumulation may limit the usefulness of such measures.

IV. SOME EXTENSIONS OF THE BASIC MODEL

The last section demonstrated the fundamental proposition that workers who decide to work rather than to shirk weigh the costs versus the benefits of shirking: the

costs of shirking come from the probability p dt of being caught and forfeiting a trust fund of earned wages which would be paid at a later date if the worker were still an employee in good standing; the benefit of shirking is v dt. Only if the trust fund is larger than v/p will it pay to work rather than shirk.

In the absence of an employment fee (which would be the worker's contribution to the trust fund at the beginning of his employment) or the firm's paying a premium (which would be the firm's initial contribution to the trust fund), it takes some time for the trust fund to build up to the point that the accrued unpaid wages exceed v/p. The amount of time required depends positively on the value of shirking to the worker and negatively on the likelihood that a shirking worker is detected by the firm. In the preceding example with e* constant, w* constant, and no discounting by either firms or workers, we showed that with no employment fees possible the firm minimized the cost of a unit of effective labor by paying into the trust fund v/p at the beginning of the worker's career and letting wages accrue at the market clearing wage rate as long as the worker continued with the firm. In this case, the worker always loses at least v/p if caught shirking and thereby chooses never to shirk. With no discounting and a constant path of w* and e* that such a policy was cheaper than paying a lower career wage bill, in which case workers would shirk early in their careers. With lower total wage payments (i.e. x less than v/p), workers would shirk for a long enough period that although total wage payments would be lower, the costs per effective labor unit would be higher than when paying a premium of v/p.

A natural question presents itself: to what extent does this result of optimal wages in excess of market clearing generalize. If there is discounting at rate r? If w* and e* vary over a worker's career?

This section will show several ways in which the preceding proposition generalizes. It generalizes if there is discounting at a constant rate r, but with e^* and w^* constant. The basic proposition holds both with and without discounting if $e^*(t)$ and $w^*(t)$ vary over time but $e^*(t)/w^*(t)$ is a constant. With general paths for $w^*(t)$ and $e^*(t)$ the firm may not pay wages in excess of market clearing. In particular, if $e^*(t)$ is increasing rapidly relative to $w^*(t)$ it will not pay the firm to pay wages in excess of market clearing for a simple reason. If $e^*(t)$ is low in the beginning relative to $w^*(t)$, the trust fund is building up rapidly relative to the amount of effective labor power the firm is losing by the worker's shirking early in his career. The firm could contribute a premium v/p and prevent shirking in the worker's early years. But, if the worker's productivity in the early years is low why should it bother?

A. Constant $w^*(t)$, Constant $e^*(t)$, and Discounting

In this subsection, we redo the analysis of section III allowing for positive discount rates by firms and workers to determine whether firms will choose to pay a present discounted value of career wages in excess of market clearing. The case of a constant outside wage w^*, a constant worker efficiency e^*, and a constant discount rate $r>0$ shared by both firms and workers is examined. As before, we assume that the firm will pay all the remuneration in the last period. A later section (section VI) examines the case of worker "myopia", where workers have a higher discount rate than firms: in that case only the premium v/p is paid in the last instant.

Suppose that the firm pays $\int_0^n w^* e^{r(n-t)} dt + (v/p)e^{rn}$ in the last period. Then the value of the trust fund is always in excess of v/p, and the worker will never

shirk. Nothing stops us from comparing this strategy of compensation to what would occur if the firm paid all compensation at retirement, but paid no premium in excess of market clearing. In that case the worker would shirk up to some date T (T>0) but not beyond. The firm would pay wages whose expected value at time 0 is

$$\int_0^n w^* e^{-rt}\, dt \tag{5}$$

and receive labor services whose discounted total would be

$$\int_T^n e^* e^{-rt}\, dt \tag{6}$$

If the ratio

$$\frac{v/p + \int_0^n w^* e^{-rt}\, dt}{\int_0^n e^{-rt}\, dt} < \frac{\int_0^n w^* e^{-rt}\, dt}{\int_T^n e^{-rt}\, dt} \tag{7}$$

then we have shown that the payment of the premium $v/p\ e^{rn}$ at time n results in cheaper labor than the payment of no premium. The LHS of (7) is the discounted cost of labor per discounted unit of effective labor with payment of the premium. The RHS of (7) is the discounted cost per unit of discounted labor in the absence of any premium. Although establishment of this equality does not show that the payment of this particular premium is the optimum, it does establish that the payment of some premium must be better than the payment of no premium at all.

As we demonstrate in Appendix 1, simple algebraic manipulation of (7) and utilization of the marginal work-shirk condition suffice to show that in fact (7) is true for all r. A slightly more subtle calculation shows that in fact payment of the premium $v/p\ e^{rt}$ at date n, a policy which totally prevents shirking, is

not only better than payment of no premium, it is in fact the optimum remuneration scheme for the firm in this setting.

B. Further Results

The preceding methods can be generalized to consider specific special cases, whose results we shall summarize here. In each case an inequality analogous to (7) is examined to see whether payments of a premium $v/p\ e^{rn}$ results in lower labor costs than market clearing wage packages. The effective labor cost for the market clearing wage is calculated using a marginal work-shirk condition.

In the following cases, it is always true that payment in excess if market clearing wages results in less costly labor.

1. $e^*(t)$ is proportional to $w^*(t)$ and there is no discounting.

2. $e^*(t)$ is proportional to $w^*(t)$, there is discounting, and both $e^*(t)$ and $w^*(t)$ are declining.

3. $e^*(t)$ is proportional to $w(t)$ but rising and the discount rate is sufficiently small.

If $e^*(t)$ is proportional to $w^*(t)$ and there is discounting, it is not in general true that payment of a premium with discounted value at the start of the job of v/p is better than paying a market clearing remuneration package.

Furthermore, if the ratio of $e^*(t)$ to $w^*(t)$ is rising, it is not always true that payments equal to the present value of the market clearing wage yield higher costs per effective labor unit than do contracts that pay a wage premium. Payment of the market clearing wage means that workers shirk until time T, but if workers are not very productive in primary sector firms early in their careers (relative

to their outside wages in the secondary sector), the firm does not lose much by not paying a premium. In the limiting case of $e^*(t)=0$ up until the marginal work-shirk time T, the firm has nothing to lose by workers' potential shirking in this interval, and there is no reason to pay wages above the market clearing level. The firms whose behavior we are modelling are high wage primary sector employers where long job tenures are prevalent and incentives for investment in skills are high. Presumably in such primary sector firms productivity grows faster than in secondary sector firms where jobs tend to be more freely available and job durations tend to be short. The finding that market clearing wages may be optimal if $e^*(t)$ rises relative to $w^*(t)$ indicates that the use of efficiency wage premiums to prevent shirking may not be a wise strategy for firms in the empirical reality we wish to model. This motivates the next section.

V. ENDOGENOUS LEARNING

The previous section showed that in the appealing case where $e^*(t)$ rises relative to $w^*(t)$ market clearing compensation packages might be optimal for primary sector employers. The attractiveness of these market clearing contracts with concomitant shirking early in a worker's career depends crucially on the growth of $e^*(t)$ not being adversely affected by this shirking behavior. It seems plausible that worker shirking will adversely affect on-the-job productivity growth. Some of the benefits of training and learning-by-doing may not arise exogenously but depend on workers paying attention to their jobs rather than shirking. In this section, we take these factors into account by assuming that $e^*(t)$ only grows in

periods during which the worker is actually working and does not rise during periods of shirking so that the worker malfeasance harms future productivity as well as current performance. When the rise in $e^*(t)$ relative to $w^*(t)$ depends on the time actually worked in the primary sector, nonmarket clearing payment schemes which provide enough of a premium to prevent shirking early in a worker's tenure are optimal. Shirking early in a worker's career is quite costly to a firm when on-the-job productivity growth is adversely affected by shirking even if the worker is not very productive during his or her early tenure. This result arises since early shirking reduces the effective labor units supplied by the worker in the primary sector later in his or her career.

A simple modification of our basic model that highlights the optimality of a wage package utilizing an efficiency wage premium in this case begins with the assumption that the effective labor supplied by a worker in a primary sector job grows exponentially at rate β during time actually worked but does not grow during periods of shirking. This implies

$$e^*(t) = e^{\beta t^*} \quad \text{where } t^* = \max(t-T, 0), \tag{8}$$

$\beta>0$, and T equals the date at which the worker's trust fund has accumulated to the point at which the worker is indifferent between working and shirking. We assume $w^*(t)$ is constant at w^* and initially assume no discounting. It is sufficient to show that the nonmarket clearing package that offers the premium v/p and thereby prevents all shirking leads to lower costs per effective labor unit than the market clearing package. We shall briefly describe the calculation which needs to be made to establish this point. The method parallels that used in previous sections.

The market clearing contract leads to shirking up to date T at which the value of the trust fund ($\int_0^T w^* dt$) times the probability of being caught shirking (p dt) just equals the gains from shirking:

$$p\,dt \int_0^T w^* dt = v\,dt. \tag{9}$$

The package involving a premium v/p is cheaper than the market clearing contract if

$$\frac{nw^*}{\int_T^n e^{\beta(t-T)} dt} > \frac{nw^* + v/p}{\int_0^n e^{\beta t}\, dt} \tag{10}$$

The numerator of the LHS is wage payments in the market clearing contract, and the denominator is the efficiency units of labor supplied by a worker who optimally decides to supply effort from T to n under this contract. The RHS analogously gives unit labor costs under a contract paying the premium v/p. Evaluation of the integrals in (10) and some algebraic manipulation suffice to show that (10) implies the inequality

$$(e^{\beta n} - 1)^2 > 0 \tag{11}$$

which is true.[13] A contract involving a premium sufficient to prevent all shirking dominates a market clearing package when $e^*(t)$ grows exponentially relative to w^* with time actually worked. This result also follows when firms and workers have a common positive discount rate. The proof in the discounting case involves considerable calculation and therefore is presented in Appendix 2.

VI. MYOPIC WORKERS

The previous solutions for optimal remuneration schedules have been derived under the assumptions that firms and workers share the same rate of discount and also are risk neutral. Because of those two assumptions, only loose restrictions are placed on the timing of payments (although the total discounted value of payments is determined). The assumption that workers have a higher rate of discount than firms (or, alternatively, have a concave utility function dependent on the current wage) will explicitly determine the wage path and thus eliminate the unrealistic feature of our previous payment schedules that all wages are paid at the date of retirement.

We assume as before that the outside wage w^* and inside labor efficiency units e^* are constant, as also are the flow return from shirking v, and the instantaneous probability of being caught p. The innovation in this section will be to assume that firms discount future payments at the rate r, while workers discount future payments at the rate δ, and with $\delta > r$ since workers on average have poorer access to capital markets than firms.

A. Optimal Remuneration Schedule

The optimal remuneration schedule has the following form when workers have a higher discount rate than firms. A worker at the end of his or her career receives a premium v/p. In every period prior to that the worker receives $w^* + \delta(v/p)$.

The preceding solution to the optimal contract can be motivated by backward induction (following the logic of Becker and Stigler (1974)). To get the worker

to work in the final instant, a premium of v/p must be paid. This premium is a fixed cost borne by the firm: the worker must have a premium of v/p in the last period (instant) worked. Now assume the worker is paid at rate $w^* + \delta(v/p)$ between time t and time n. These payments are just at the minimum level so that the worker at each instant has a surplus of v/p if employed, and therefore chooses to work rather than shirk. Then ask the question how much the worker must be paid to work rather than shirk from t-dt to t. This could be done by a payment

$$(w^* + \delta\ (v/p))\ dt$$

between t-dt and t. The firm could also make higher payments at a later date and reduce the payment between t-dt and t. Yet, these higher future payments would have to be more expensive to the firm, whose discount rate r is lower than the worker's discount rate δ. Thus the optimal schedule to keep the worker form shirking is to pay at a rate $w^* + \delta(v/p)$ from 0 to n. Appendix 3 shows rigorously that the minimum cost contract to keep the worker working from 0 to n is to pay a constant wage $w^* + \delta(v/p)$ plus a final payment v/p at the retirement date.

It is also true that it pays the firm to induce the worker not to shirk throughout the worker's whole career. A payment v/p represents a fixed cost to create work incentives for a given worker. In the worker's last instant of working, there must be a surplus of at least v/p. If this surplus is at its minimum, wages paid discounted at rate r must be minimized by paying at $w^* + \delta(v/p)$ for the worker's whole career. The firm receives the lowest possible labor cost if it spreads the fixed cost v/p over the longest possible working career. Thus the optimum remuneration schedule is to pay at the rate $w^* + \delta(v/p)$ from 0 to n with a final payment of v/p.

B. Remarks

Five remarks are in order about optimal remuneration with worker myopia. First, this solution is in agreement with the more general solutions found by Kuhn (1986). Kuhn, however, did not discuss the extent to which remuneration schedules give a surplus to workers above the market clearing level.

As a second remark, it is of interest to consider how the results are modified if workers cannot borrow, have a discount rate δ (not necessarilly equal to r), and have an instantaneous concave utility function U(C) where C is consumption. The more interesting case here is one in which workers live for a significant period after retirement. If the post-retirement period is sufficiently long that the use of a contingent pension (or retirement annuity) of value v/p imposes no upward jump in consumption at retirement, then the optimal contract pays a constant wage premium over the entire worklife plus a retirement annuity of sufficient value to prevent shirking late in a worker's career. In such a case, the worker's present discounted value of retirement consumption will be at least v/p.[14] The consideration of worker risk aversion does not alter the conclusion that optimal labor contracts are nonmarket clearing when there is a shirking problem and upfront bonds are not feasible.

As a third point, it is worth noting two numerical examples for our case of risk neutral workers. Fixing w^* at 1, if v=.2, δ=.1 per year, and p=.2 per year (a number which might be realistic in cases where shirking is hard to detect), the optimal wage w will be 10 percent greater than the outside wage w^* and the final premium will be equal to one year's outside wages. Thus wages significantly in excess of w^* occur in this example. The probability of being caught shirking (or, to be more precise, the probability of sufficient evidence of malfeasance having been uncovered that the firm can take disciplinary action) in this example is

rather low. If shirkers can be detected easily and also easily dismissed if detected, wage premiums are not likely to be very large. Consider the same parameters ($w^*=1$, $\delta=.1$ per year, $v=.2$) with a change in p so that worker has a .5 chance of being dismissed for shirking in a month. Translating p into an annual basis, one finds that the optimum wage w is 1.0024 and the retirement premium is .024. So cases where detection is fairly probable over the course of a month result in small premiums.

The preceding observation may help explain the nature of differences between primary and secondary labor markets. Doeringer and Piore (1971) note that most jobs can be classified into those which have close monitoring, rapid dismissal for worker malfeasance and low wages and those with loose monitoring, slow dismissal for fault, and high wages. If jobs differ according to p, the technology of monitoring, such a pattern is explained.[15]

As a fourth observation on the myopia wage paths, it is worth remarking that even the smallest amount of myopia by workers (i.e. the smallest difference between δ and r) leads to a determinate wage schedule.

Finally, we note that our model suggests that Lazear's (1979, 1981) upward sloping wage schedules prior to retirement are not the direct result of implicit bonding but instead must arise from an upward sloping w^* schedule and/or the absence of reputation by firms and workers, so that both may cheat. In our model where w^* is constant and the threat of lost reputation prevents firms from cheating, the wage schedule is flat until retirement. Beyond a lump sum payment at the retirement date, an increasing wage schedule in our model does not occur as an inducement to get workers not to cheat.

VII. ENDOGENOUS MONITORING

In the previous models, workers have greater incentive to shirk early in their careers than late in their careers because the trust funds placed at risk when workers shirk grow with job tenure. Thus firms should know that the prevention of shirking requires more intensive monitoring of workers early in their careers than late in their careers. The previous assumption of an exogenous and constant instantaneous probability of being caught shirking is therefore unrealistic. An endogenous choice of monitoring intensity by firms would be quite likely to generate a probability of being caught shirking that declined with workers' job tenure. The previous analysis has not answered whether such a pattern for p(t) would alter the conclusion that primary sector firms will choose to pay wage premiums. This section deals with that essential question. It yields the conclusion that wage premiums will be paid in a model where firms optimally choose monitoring intensity according to worker's life cycle pattern of temptations to shirk.

This analysis of endogenous monitoring requires the specification of a monitoring technology. The probability of detecting a shirker is assumed to be proportional to monitoring expenditures (i.e. expenditures on shirking detection.) The optimal monitoring strategy turns out to involve monitoring a worker intensively at the beginning of his or her career with the level of monitoring expenditures declining with job tenure as his or her trust fund grows. A wage premium above the market clearing level turns out to be optimal. The wage premium depends positively on the value of shirking and the costs of monitoring.

We return to the analysis of the basic model of section III (constant e^*, constant w^*, and no discounting) with the single modification that the firm has a choice over expenditures on shirking detection. The firm seeks to pick a remuneration package and monitoring strategy to minimize unit labor costs. A compensation package in this no myopia case consists of a level of total payments made to the worker at the end of his or her career contingent on the worker having never been caught shirking. A monitoring strategy involves a choice of monitoring expenditures and thereby a shirking detection probability at each point of time in a worker's career. The probability of detecting a worker shirking in the small interval t to t+dt, denoted p(t) dt, is proportional to expenditures made in monitoring the worker over the interval. This implies the monitoring costs over the interval t to t+dt required to generate a shirking detection probability p(t) dt are equal to cp(t) dt where c is a constant.

The problem for the firm is to pick a wage premium x over the market clearing package to be paid at the end of the workers career and a level of monitoring expenditures at each date. The choice of a level of monitoring expenditures at each date is equivalent to a choice of the path for p(t). If the firm finds it optimal at some date not to spend enough on monitoring to prevent the worker from shirking, then the firm will set monitoring expenditures to 0 at that date. The firm's problem is to choose x and the path of p(t) to minimize

$$(12) \qquad (1/(n-T))[x + w^*n + \int_T^n cp(t)\, dt]$$

The term in brackets in (12) is the sum of compensation costs ($x+w^*n$) and monitoring costs ($\int_T^n cp(t)\, dt$). T is the date at which the worker's trust fund is just

large enough that he or she is on the margin between working and shirking. If T is greater than 0, then p(t) equals 0 for t less than T.

The problem is simplified by noting that the no shirking condition at date t is given by

(13) $$p(t)\, dt(tw^* + x) = v\, dt$$

where tw^*+x is the trust fund accumulated up to t. This no shirking condition implies a time pattern for p(t):

(14) $$p(t) = v/(tw^*+x).$$

Thus once the no shirking condition holds, monitoring intensity and the likelihood of shirking detection declines over the course of a worker's career as the temptation to shirk ($v/[tw^*+x]$) falls.

The substitution of the no shirking condition (14) into (12) and integration yields a function for unit labor costs

(15) $$C(x,T) = \frac{x + nw^* + \frac{cv}{w^*} \ln\left(\frac{nw^*+x}{Tw^*+x}\right)}{n - T}$$

The solution of the firm's problem involves minimizing C(x,T) with respect to x and T. Consideration of the function shows an optimum of T=0 with a premium x>0. (The argument is just a bit complicated. Substituting the first order condition for x into the first order condition for T shows that $\partial C/\partial T>0$. With T=0, $\partial C/\partial x<0$ for x=0.)

The analysis of this endogenous monitoring model has also been extended to the case of a positive discount rate which is common to both workers and firms. The derivations are quite messy and do not lead to an expression for $\partial C/\partial T$ that can

be signed analytically. We have numerically determined the optimal solution for a wide variety of choices of the parameters v, c, n, and the discount rate r. In each case analyzed, the optimal solution involved T=0 (the no shirking constraint holding at all points of time) and a positive wage premium above nw*.

Remark:

Since monitoring and close supervision of workers are costly, firms should make workers post very large bonds and spend little or nothing on monitoring. The fact that firms do devote considerable resources to monitoring employees for the purpose of controlling worker malfeasance is direct evidence that bonding is limited and that it is costly on the margin.[16] Thus it should not be surprising in a world where the monitoring of workers accounts for a substantial portion of labor costs that firms also utilize efficiency wages to provide incentives for adequate employee performance.

VIII. CONCLUSION

This paper has shown that the bonding critique justifies no a priori presumption of an absence of efficiency wage premiums. Explicit upfront bonds are only rarely observed. In the absence of such bonds, upward sloping wage profiles do not act as a perfect substitute in providing incentives not to shirk. On the contrary, in the absence of such explicit bonds, our paper suggests that second best optimal remuneration schedules will be in excess of market clearing where shirking is a potential problem and monitoring is difficult. The existence or nonexistence of

an important role for efficiency wages is essentially an empirical, not a theoretical, problem.

The findings of this paper suggest the need for empirical work on the importance of efficiency wages. First, research is needed to document the existence or the nonexistence of sizeable and persistent wage premiums above the market clearing level to see if the existing wage structure is compatible with the predictions of a textbook competitive labor market model. If wages differ substantially from those that can be reasonably justified by the standard competitive model, do they systematically differ in the ways predicted by efficiency wage theory? Recent work by Dickens and Katz (1986a,b) and Krueger and Summers (1986a,b) has demonstrated the existence of substantial, persistent interindustry wage differentials for "similar" workers (i.e. those that are "observationally equivalent" to the econometrician).[17] If there is a distribution of wages for similar workers, some workers must be earning wage premiums.

There remains the question of stigma and the importance of worker reputations on labor market outcomes. Is the stigma from dismissal for cause from a job a sufficient discipline device that wage premiums are not necessary? The importance of such stigma constitutes an important question for future empirical research.

FOOTNOTES

1. Models of this type have recently been examined by Bowles (1985), Bulow and Summers (1986), Calvo (1985), Eaton and White (1983), Foster and Wan (1984), Gintis and Ishikawa (1983), Jones (1985), Shapiro and Stiglitz (1984), and Stoft (1982). Stiglitz (1984) and Yellen (1984) provide excellent surveys of alternative efficiency wage models of unemployment. Katz (1986) surveys recent theoretical developments and empirical work related to efficiency wage models.

2. Furthermore, it has been noted by Carmichael (1985) that this property does not depend on workers' ownership of assets if workers are sufficiently risk averse or access to financial markets: the amount of the bond will be bid up in equilibrium to the point where the marginal worker is indifferent between taking a job, which pays a wage but requires an initial payment to the firm by the worker, and being unemployed. No unemployed worker would prefer to work at the prevailing wage and posting the bond for employment; capital market imperfections cause the required bonds to be small, but will not cause the nonclearance of the labor market. So unemployment is voluntary in this case. If, however, workers are close to risk neutral, and have very limited access to capital markets, putting at risk the small amount of bond they are willing to offer may not be enough to induce them not to shirk. The generality of Carmichael's point depends on instantaneous utility going to $-\infty$ if workers pay a firm a bond of all their "liquid" assets.

3. MacLeod and Malcomson (1986) make a somewhat different assumption within the context of a shirking model. They assume that firms can observe shirking but that shirking is not public information in the sense that firms have a hard time proving that a worker has shirked to a third-party such as a court or arbitrator. This reflects the European tradition of requiring just cause for dismissal. Recent court interpretations and legislation in several states suggest the United States is moving in this direction away from a strict employment-at-will doctrine. Still, the examples cited in the text indicate that the actual unobservability of worker malfeasance by employers may be important regardless of whether observed shirking is public information or not.

4. See Carmichael (1985) for an argument why involuntary unemployment may not result even when firms cannot be trusted not to falsely default and Shapiro and Stiglitz (1985) for several objections to his argument.

5. See Lazear (1981) for an emphatic statement of this point.

6. This folk argument is asserted in discussions of the bonding critique of efficiency wage models in Akerlof and Yellen (1984) and Yellen (1984) as well as implied in many other papers and in many discussions at conferences which we have attended.

7. This holds provided shirkers must be fired when caught rather than at a later date. The reasons for and the implications of this assumption are discussed in footnote 9 below.

8. Market clearing wages may provide sufficient incentive to prevent workers from shirking throughout their careers if there are substantial costs associated with the stigma of being fired for shirking or if their are substantial costs to moving across jobs.

9. This assumption is plausible if the detection of shirking by the firm is observable to the worker, since the worker would continue shirking or quit if he or she knew that he or she would get sacked before receiving deferred payments. Alternatively, the delayed informing of the worker that he or she has been caught shirking and in danger of disciplinary action is likely to leave a firm open to an unjust dismissal suit for violation of the covenant of good faith in many states in the United States and would not be permissible under the dismissal rules in most European economies.

10. We assume that firms' production functions are of the form $f(e^*n)$ where n is the number of laborers who are supplying effort e^*. A worker who supplies 0 labor has no effect on output. It is said that a worker who shirks supplies 0 units of effective labor while a worker who works supplies e^* units of effective labor.

11. If the firm could hide its knowledge of having detected a worker shirking and wait till n before dismissing a worker for a shirking offense committed at t, the worker's potential loss from shirking at t is instead

$$p\ dt\ (w^*n+x)$$

since a worker dismissed an n will attain no outside earnings. In this case, the firm's optimal strategy is to hide its knowledge of having caught a worker shirking and wait till just prior to till just prior to the worker's retirement date to fire the worker. If the worker knows that he or she has been caught shirking this strategy has no use. At that point, the worker will seek other employment. Also, such behavior by firms may leave them open to unjust dismissal suits or union grievances (see footnote 9).

12. Hutchens (1986) shows in a shirking model in which workers are assumed to be able to post upfront bonds that the specter of firm cheating on delayed payments introduces a form of fixed costs into the employment relationship. Since a firm entails these fixed costs each time it hires a new worker, firms prefer to hire young workers with long potential tenures. Hutchens argues that firms with reputations for honesty do not face these fixed costs and should be indifferent between hiring young and old workers. In contrast, our model with limits to upfront bonding shows that even honest firms face the fixed costs of generating enough surplus to provide work incentives. If upfront bonds are limited, firms will prefer to hire young workers so that the fixed premium required to induce workers not to cheat is amortized over the longest period possible. This preference for workers with long potential careers follows whether or not firms may potentially cheat on their deferred payments to workers.

13. Evaluation of the integrals in (10) and a little algebra directly generate the inequality

$$(n/T)(e^{\beta n}-e^{\beta(n-T)}) > e^{\beta(n-T)}-1.$$

This can be weakened to

$$e^{\beta n}-2e^{\beta(n-T)}+1 > 0.$$

This inequality produces the inequality noted in the text since

$$e^{\beta n} - 2e^{\beta(n-T)} + 1 \geq (e^{\beta n} - 1)^2 > 0.$$

14. There is a second case where payments of value v/p at retirement severely distort a worker's consumption profile. For example, it would entail a large upward blip in consumption during the retirement period if workers live for only a short time after the retirement date. Such a jump in consumption is quite costly to risk averse workers. If risk aversion is strong enough and the losses from shirking are small enough, it may pay the firm to spread out the payment of the surplus v/p over a discrete interval immediately before the retirement date rather than paying it as a lump sum at retirement. This type of contract will induce workers to shirk very late in their careers. Still, the reduction in total worklife payments may be large enough to offset the losses from shirking by workers close to retirement.

15. This observation is the basis for formal models of dual labor markets developed by Bulow and Summers (1986) and Jones (1985).

16. See Dickens, Katz, and Lang (1986) and Carmichael (1986) for more detailed discussions of this point.

17. See Summers (1986) for a discussion of some of the key conceptual issues in trying to identify efficiency wage premiums.

REFERENCES

Akerlof, George A. and Janet L. Yellen. "Efficiency Wage Models of the Labor Market: Introduction." Mimeo, U.C. Berkeley, 1984.

Allen, Steven G., Robert L. Clark, and Daniel A. Sumner. "Post-Retirement Adjustments of Pension Benefits." NBER Working Paper No. 1364, June 1984.

Axelrod, Robert. The Evolution of Cooperation. New York: Basic Books, 1984.

Becker, Gary S. and George J. Stigler. "Law Enforcement, Malfeasance, and the Compensation of Enforcers." Journal of Legal Studies 3 (January 1974): 1-18.

Bowles, Samuel."The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian and Marxian Models." American Economic Review 75 (March 1985): 16-36.

Bulow, Jeremy I. and Lawrence H. Summers. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination and Keynesian Unemployment." Journal of Labor Economics 4 (July 1986): 376-414.

Burawoy, Michael. Manufacturing Consent. Chicago: University of Chicago Press, 1979.

Calvo, Guillermo. "The Inefficiency of Unemployment: The Supervision Perspective." Quarterly Journal of Economics 100 (May 1985): 373-87.

Carmichael, H. Lorne. "Can Unemployment Be Involuntary?: Comment." American Economic Review 75 (December 1985): 1213-14.

----------. "Efficiency Wages and Labor Hoarding: A Survey." Mimeo, Queen's University, June 1986.

Dickens, William T., Lawrence F. Katz, and Kevin Lang. "Are Efficiency Wages Efficient?." NBER Working Paper No. 1935, June 1986.

Dickens, William T. and Lawrence F. Katz. "Inter-industry Wage Differences and Theories of Wage Determination." Mimeo, U.C. Berkeley, revised July 1986a.

----------. "Interindustry Wage Differences and Industry Characteristics." In Kevin Lang and Jonathan Leonard, eds., Unemployment and the Structure of Labor Markets. Oxford: Basil Blackwell, forthcoming 1986b.

Doeringer, Peter B. and Michael J. Piore. Internal Labor Markets and Manpower Analysis Lexington, Ma.: D.C. Heath, 1971.

Eaton, Curtis and William D. White. "The Economy of High Wages: An Agency Problem." Economica (April 1983): 175-81.

Foster, James E. and Henry Y. Wan. "Involuntary Unemployment as a Principal-Agent Equilibrium." American Economic Review 74 (June 1984): 476-84.

Gintis, Herbert and Tsuneo Ishikawa. "Wages, Work Discipline and Macroeconomic Equilibrium." Mimeo, 1983.

Hutchens, Robert. "Delayed Payment Contracts and a Firm's Propensity to Hire Older Workers." Journal of Labor Economics, forthcoming 1986.

Jones, Stephen R.G. "Dual Labor Markets, Productivity and Unemployment." Mimeo, University of British Columbia, April 1985.

Katz, Lawrence F. "Efficiency Wage Theories: A Partial Evaluation." In S. Fischer, ed., NBER Macroeconomics Annual 1986. Cambridge, Ma.: MIT Press, forthcoming 1986.

Krueger, Alan and Lawrence H. Summers. "Efficiency Wages and the Wage Structure." Mimeo, Harvard University, May 1986a.

----------. "Reflections on the Inter-industry Wage Structure." In K. Lang and J. Leonard, eds., Unemployment and the Structure of Labor Markets. Oxford: Basil Blackwell, forthcoming 1986b.

Kuhn, Peter. "Wages, Effort, and Incentive Compatibility in Life-Cycle Employment Contracts." Journal of Labor Economics 4 (January 1986): 28-49.

Lazear, Edward. "Agency, Earnings Profiles, Productivity, and Hours Restrictions." American Economic Review 71 (September 1981): 606-620.

----------. "Why is there Mandatory Retirement?" Journal of Political Economy 87 (December 1979): 261-84.

Macaulay, Stewart. "Non-contractual Relations in Business: A Preliminary Study." American Sociological Review 28 (February 1963): 55-67.

MacLeod, W. Bentley and James M. Malcomson. "Implicit Contracts, Incentive Compatibility, and Involuntary Unemployment." Queen's University Discussion Paper No. 585, March 1986.

Mars, Gerald. Cheats At Work. London: Allen and Unwin, 1982.

Roy, Donald. "Restriction of Output in a Piecework Machine Shop." Unpublished doctoral dissertation, University of Chicago, 1952.

Shapiro, Carl and Joseph E. Stiglitz. "Equilibrium Unemployment as a Worker Discipline Device." American Economic Review 74 (June 1984): 433-44.

----------. "Can Unemployment Be Involuntary?: Reply." American Economic Review 75 (December 1985): 1215-17.

Stiglitz, Joseph E. "Theories of Wage Rigidity." NBER Working Paper No. 1442, September 1984.

Stoft, Steven. "Cheat Threat Theory: An Explanation of Involuntary Unemployment," Mimeo, Boston University, May 1982.

Summers, Lawrence H. "Did Henry Ford Pay Efficiency Wages?." Mimeo, Harvard University, April 1986.

Yellen, Janet L. "Efficiency Wage Models of Unemployment." American Economic Review Proceedings 74 (May 1984): 200-205.

APPENDIX 1

PROOF THAT PAYMENT OF PREMIUM $(v/p)e^{rn}$ AT TIME n RESULTS IN LOWER UNIT LABOR COSTS THAN PAYMENT OF NO PREMIUM (WITH CONSTANT w* AND e* AND DISCOUNTING AT RATE r)

The proof follows by deriving the exact formula for T. The worker can view the deferment of his payment until n in the market clearing case as the building up of a trust fund whose value at time t, with discounted payments exactly matching those in the secondary sector, is given by

(A1-1) $$\int_0^T e^{r(t-\tau)}\, w^*\, d\tau.$$

And the worker is just indifferent between working and shirking at time T where the expected loss from shirking, which is

(A1-2) $$p\, dt \int_0^T e^{r(T-\tau)} w^*\, d\tau$$

is equal to the gain from shirking, which is v dt. In other words, the time at which the worker ceases to shirk is the time when the accumulated value of the trust fund to be forfeited if caught shirking has just reached v/p. Thus, T satisfies

(A1-3) $$w^* \int_0^T e^{r(T-t)}\, dt = v/p.$$

Rewriting the denominator on the RHS of (7) as

(A1-4) $$\int_0^n e^{-rt}\, dt - \int_0^T e^{-rt}\, dt,$$

noting that

(A1-5) $$\int_0^T e^{-rt}\, dt = [(v/p)/w^*]e^{-rT},$$

and algebraic manipulation of (7) shows that this key inequality will hold if and only if the following inequality holds:

(A1-6) $$\int_0^n e^{-rt} dt - \int_0^T e^{-rt} dt - \left(\int_0^n e^{-rt} dt \right) e^{-rT} < 0.$$

Evaluation of the three integrals in (A1-6) shows the RHS to be

(A1-7) $$[e^{-rn}/r]\,(e^{-rT}-1),$$

so the inequality (A1-7) and hence inequality (7) always hold. With discounting, as without discounting, the firm pays less per discounted efficiency unit of labor if it pays a premium above the wages in the secondary labor market.

APPENDIX 2

PROOF THAT NONMARKET CLEARING WAGES ARE OPTIMAL IN THE CASE OF ENDOGENOUS LEARNING AND DISCOUNTING

The marginal no shirking condition is given by

$$p\,dt \int_0^T e^{-rt} w^* dt = v\,dt. \tag{A2-1}$$

If $\beta=0$ and $r>0$, we know that

$$\frac{\int_0^n e^{-rt}\, w^*\, dt}{\int_T^n e^{-rt} e^{\beta(t-T)} dt} > \frac{\int_0^n e^{-rt}\, w^*\, dt + v/p}{\int_0^n e^{-rt}\, e^{\beta t} dt} \tag{A2-2}$$

since we proved in Appendix 1 that in this case the payment of a premium of v/p leads to cheaper labor than the market clearing package. The RHS of (A2-2) is average discounted labor costs at market clearing wages. The LHS of (A2-2) is average discounted labor costs paying the premium v/p. So the question becomes one of showing that

$$\frac{d}{d\beta}\ (D/B) \geq 0$$

where $D = \int_T^n e^{-rt} e^{\beta(t-T)} dt$ and $B = \int_0^n e^{-rt} e^{\beta t} dt$.

Evaluation of the integrals D and B and taking the derivative of (D/B) with respect to β shows it to be proportional to

$$f(y) = (1-\alpha)e^{(1+\alpha)y} - e^{y} + \alpha e^{\alpha y}$$

with a positive constant of proportionality and where $y=n(\beta-r)$ and $\alpha=(n-T)/n$. We analyze $f(y)$ in two cases.

Case I: $y>0$ $(\beta>r)$

Taking successive derivatives of y, one finds that the Taylor's expansion of f(y) around y=0 is given by

$$f(y) = \sum_{m=2}^{\infty} ((1-\alpha)(1+\alpha)^m - 1 + \alpha^{m+1}) y^m / m!$$

$$= \sum_{m=2}^{\infty} \left[\sum_{j=1}^{m} \{ \binom{m}{m-j} - \binom{m}{m-j+1} \} \right] \frac{y^m}{m!} \quad .$$

Each term in square brackets in the above expression is positive. Higher ordered terms of α^j are negative, while lower ordered terms are positive with equal coefficients. The symmetry of the binomial expansion allows higher order and lower order terms in α to be matched up in such a manner that each term in square brackets is easily demonstrated to be positive. Thus, $f(y)>0$ for $y>0$.

Case II: $y<0$ ($\beta<r$)

It remains to show that f(y) is positive for y between $-\infty$ and 0. This can be accomplished by first considering the function g(z) where

$$g(z) = (1-\alpha)z^{(1+\alpha)} - z + \alpha z^{\alpha}$$

and $z=e^y$. Showing $g(z)>0$ for $0<z<1$ is equivalent to showing $f(y)>0$ for $y<0$. One can easily show that $g(1)=0$, $g'(1)=0$, $g''(1)>0$, $g(0)=0$, and that g(z) approaches 0 from above as $z \to 0$. An evaluation of g''(z) shows that g has only one inflection point between 0 and 1. With only one inflection point in the range 0 to 1, the properties of g noted above imply g cannot have a root between 0 and 1. g is greater than 0 for $z>0$ and $z<1$. If g is ever less than 0 for z between 0 and 1 it must at some point fall below 0. If g falls below 0 in this interval it must have an interior local minimum as well as two interior local maxima. This cannot occur with only one inflection point. Since g has no roots between 0 and 1, f(y) is positive for $y<0$.

The proof that the payment of a premium of discounted value v/p dominates the market clearing contract in the case of $\beta=r>0$ follows directly from a comparison of the average unit labor costs in each case and some tedious algebraic manipulation.

APPENDIX 3

DERIVATION OF MINIMUM COST COMPENSATION PACKAGE WITH WORKER WORKER MYOPIA

In this appendix, we show that if a firm makes workers a final payment of v/p at the retirement date n, the wage package which has the lowest present discounted value to the firm and which prevents shirking throughout a worker's career will be the payment of wages at constant rate $w^*+\delta(v/p)$ fron 0 to n.

The firm wishes to minimize

$$\int_0^n w(t)e^{-rt}dt$$

subject to the no shirking constraint at each point in time, which is

$$\int_t^n (w(\tau)-w^*(\tau))e^{-\delta(\tau-t)}d\tau + e^{-\delta(n-t)}v/p \geq v/p.$$

Define in canonical notation the variable x, where

$$x = \int_t^n e^{-\delta\tau}w(\tau)d\tau.$$

Then $$\dot{x} = -e^{-\delta t}w(t)$$

and the firm wishes to maximize

$$\int_0^n \dot{x}e^{(\delta-r)t}dt$$

subject to the constraint $x(t)\geq f(t)$

where $$f(t) = [(w^*/\delta) + (v/p)](e^{-\delta t}-e^{-\delta n}).$$

This maximization problem results in the LaGrangian expression

$$\mathcal{L}\ (x,\dot{x},t,\lambda) = \dot{x}e^{(\delta-r)t} + \lambda(x-f).$$

The Euler equation

$$\frac{\partial \mathcal{L}}{\partial x} - \frac{d}{dt}\left(\frac{\partial \mathcal{L}}{\partial \dot{x}}\right) = 0$$

yields

$$\lambda = e^{(\delta - r)t} > 0$$

Thus $x(t)=f(t)$ for all t. As a result

$$-e^{-\delta t}w(t) = \dot{x}(t) = -e^{-\delta t}(w^* + \delta(v/p))$$

and

$$w(t) = w^* + \delta(v/p).$$

Hence the optimal payments schedule which prevents the worker from shirking from 0 to n and makes the final fixed payment v/p is

$$w(t) = w^* + \delta(v/p).$$

for t from 0 to n.