NBER WORKING PAPER SERIES

MALFEASANCE IN LONG TERM EMPLOYMENT CONTRACTS: A NEW GENERAL MODEL WITH AN APPLICATION TO UNIONISM

Peter Kuhn

Working Paper No. 1045

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge MA 02138

December 1982

This research was supported by a grant from the Sloan Foundation. The research reported here is part of the NBER's research program in Labor Studies. Any opinions expressed are those of the author and not those of the National Bureau of Economic Research.

NBER Working Paper #1045
December 1982

Malfeasance in Long Term Employment Contracts: A New General Model with an Application to Unionism

ABSTRACT

This paper argues that the structure of long-term employment contracts is influenced by the possibility that at least four different kinds of opportunistic behavior, or "malfeasance," may occur in them. While the consequences of some of these problems have been examined in various papers, no single model has yet treated all four and thus brought out their essential symmetry. In particular, a certain kind of malfeasance by firms has apparently been universally overlooked -- an oversight we try to remedy by developing a simple model here. Other advantages of the present model are that, unlike other models, it endogenizes the path of <u>both</u> sides of the contract -- wages and effort -- and has fairly intuitive first-order conditions. It also shows how earlier conclusions, such as the notion that wages are likely to rise faster than marginal products in equilibrium, are the results of less-than-general model specification, and has some interesting implications when applied to unionism: by proposing that unions act as workers' equivalent to certain contract enforcement policies like the disciplinary dismissals used by firms, it provides what is to the author's knowledge the only consistent theoretical explanation of the quite commonly observed U-shaped pattern of the union wage effect by age and shows how unions might play a positive efficiency role in this regard.

Peter Kuhn
Department of Economics
Harvard University
Cambridge, MA 02138

I. INTRODUCTION

This paper argues that the structure of long-term employment contracts is influenced by the possibility that at least four different kinds of opportunistic behavior, or "malfeasance," may occur in them. While the consequences of some of these problems have been examined in various papers, no single model has yet treated all four and thus brought out their essential symmetry. In particular, a certain kind of malfeasance by firms has apparently been universally overlooked -- an oversight we try to remedy by developing a simple model here. Other advantages of the present model are that, unlike other models, it endogenizes the path of <u>both</u> sides of the contract -- wages and effort -- and has fairly intuitive first-order conditions. It also shows how earlier conclusions, such as the notion that wages are likely to rise faster than marginal products in equilibrium, are the results of less-than-general model specification, and has some interesting implications when applied to unionism: by proposing that unions act as workers' equivalent to certain contract enforcement policies like the disciplinary dismissals used by firms, it provides what is to the author's knowledge the only consistent theoretical explanation of the quite commonly observed U-shaped pattern of the union wage effect by age and shows how unions might play a positive efficiency role in this regard.

The paper begins in Section II with an overall discussion of the malfeasance issue and a brief review of the existing literature in light of that discussion. The model's structure is set out in Section III, and Section IV draws out some of its more basic implications. The model is actually solved for optimal wage and effort profiles in Section V, while

Section VI applies the model more directly to the question of unions' effects on compensation profiles by solving an example. Conclusions are summarized in Section VII.

II. REVIEW: THE GENERAL PROBLEM OF MALFEASANCE IN LONG-TERM EMPLOYMENT CONTRACTS

In general, any contract between a firm and a worker over a significant period of time T, involves an agreement on both what the firm supplies to a worker -- a complex bundle of wages, fringes, and working conditions which we assume is summarized by the measure $w(t)$ -- and what the worker supplies in return, which we might loosely term "effort" or $h(t)$. Any such contracts can in principle be broken by both parties in two quite distinct ways -- first, either party has the option, at any time t, of unilaterally terminating the contract by withdrawing from it, and second, either party may attempt to alter the initial terms of the contract *ex post*, by supplying less of the factor it supplies than was (implicitly) agreed upon, or perhaps by making an all-or-nothing demand for more of the other party's factor -- and the possibility that any of the above will occur is a crucial factor determining the structure such contracts will have.

In short, in designing long run contracts, firms and workers must take account of the possibility of four different problems arising during the life of the contract: First, workers may quit the firm in order to avoid their contractual obligations to the firm if they can do better elsewhere. Second, firms may dismiss workers to whom they owe more than they expect to receive for the same reason. We shall refer to these two activities henceforth as "unilateral withdrawal" from the contract, or "Type 1" malfeasance -- even though an activity like worker quitting may not often be thought of in these terms. Third, workers at some point during the contract may all of a sudden decide to supply less effort than was initially agreed upon, or to raise their wages by stealing or other means. Finally, the firm at any point might begin to let working conditions deteriorate, fail to increase wages at the agreed-upon rate, or demand speed-ups or stretch-outs not expected when the "contract" was made. These last two activities might be called "malfeasance by trying to alter the terms of the contract," or simply "Type 2" malfeasance.

Unfortunately, even though all the above problems can at least potentially impose serious constraints on the structure of contracts, they are rarely if ever all considered together in the literature.

The earliest and simplest implicit contracts literature, like Baily (1974) and Azariadis (1975) for example, tended not to consider any of the above problems very closely, describing optimal wage and employment policies which were binding on both firms and workers in subsequent periods, invoking firm reputation and arguing that workers were locked into firms by job-changing costs to justify this procedure. The realization that worker's ability to quit, however, might impose important constraints on such employment contracts (since for example firms would have difficulties collecting "premiums" in providing workers insurance against cyclical fluctuations), was soon incorporated into formal models such as Grossman (1978), Holmstrom (1980) and Harris and Holmstrom (1981) all of whom derived various implications of what we call Type 1 malfeasance by workers for contract structure. These models however still relied on "reputation" to rule out the same kind of cheating by firms, which appears not to have been formally treated until Hashimoto and Yu (1980) and Hall and Lazear (1982), who consider symmetrical models in which the wage/effort bargain is prespecified but both parties to the contract have the right to terminate it unilaterally. Again, "second best" contract structure, plus the result that quits and dismissals are greater than optimal, are derived. Finally, a number of models of contracts that deal primarily with the problem of eliciting effort from workers (Type II malfeasance) have been developed, such as Eaton and Rosen (1981) or Lazear and Rosen (1981). These tend not to deal with any sort of firm malfeasance at all, except for the interesting case of Lazear (1979, 1981): in this model, workers cheat by changing the terms of the contract (Type II), i.e., supplying less effort in "shirking", or stealing, while firms cheat by unilaterally terminating the contract (Type I). Some mention of worker quits is made (1981, p. 612, the reference to "slavery") but it is not treated formally, and Type II cheating by firms is not considered.

The present model uses some of the notation and mechanics of Lazear's model , thus dealing with the same kind of long-run contracts (where the issue is primarily allocation of income over the life cycle rather than over states of nature as in Azariadis (1975), for example), albeit in a very simple case where there is no uncertainty. While the need for this abstraction is unfortunate, it allows us to increase the generality of the model in other directions and make it truly symmetrical while keeping it tractable, and even helps to make the basic intuition regarding how malfeasance affects contract structure more apparent.

III. MODEL STRUCTURE

The basic spirit of our model here can be outlined quite succinctly. We are in a world of complete certainty and workers are identical. Workers join a firm for a fixed period of time T which includes the retirement period and in doing so implicitly agree on a contract stipulating the path of total compensation paid to workers, w(t), and total labor supply, including "effort", h(t). Both w(t) and h(t) can be observed by either party, with the stipulation that there is a lag before the party who does not initiate a change observes that they have diverged from their agreed-upon path. The labor market is competitive, with a large number of (potential) firms willing to offer an "optimal" contract to a worker of any age, which pays the worker the present value of his or her marginal product, but immobilities of factors are allowed to exist (indeed we shall find them essential) due to hiring costs, search costs for new jobs, and/or specific human capital.

In this world, there are three main sets of factors influencing the shape of w(t) and h(t): First, since workers' productivity and valuation of effort changes over time, and since effort produces diminishing marginal returns and increasing marginal disutility, there is some way to allocate h optimally over time. Second, since there is progressive taxation and an imperfect capital market, workers will also not be indifferent between w(t) streams of equal present value, so there is some incentive to allocate w optimally over time. Third, at all points during the contract, there is a certain amount both workers and firms can gain by breaking that contract in the most profitable way, which is in most cases positively related to how well they could do by leaving the firm and employing "their" factor elsewhere, although as we shall see, it varies with the kind of malfeasance contemplated. The contract (w(t), h(t)) must be such that the *ex post* expected present value of profits and utility are high enough to make malfeasance unattractive to both parties, and the harder it is for one party to the contract to discipline the other, the more limited the range of w(t) and h(t) that do not lead to cheating will be, and the more the contract's ability to allocate w(t) and h(t) over the life cycle optimally will be compromised. As we shall see, this has interesting applications to the question of unions and life-cycle compensation and effort profiles.

More formally, the structure of the model is outlined below in three stages. First, the maximand for the optimal contract and the zero-profit constraint are introduced. Second, the constraints that all four kinds of malfeasance place on the structure of the optimal contract are developed in turn, and finally the competitive labor market context of the model is discussed.

1. Utility and Profits in the Optimal Contract

Workers' instantaneous utility at time t in our model is assumed to be given by

$$W(t) = \{U(w(t)) - \frac{1}{a(t)} V(h(t))\} e^{-\delta t} \tag{1}$$

$$U'>0,\ U''<0;\quad V'>0,\ V''>0$$

where δ is the workers' subjective discount rate, U is utility of compensation and can be thought of as embodying the effects of both imperfect capital markets and progressive taxation on the desire to smooth income, V is disutility of effort, and a(t) is a function allowing workers' preferences for leisure to vary over the life cycle (e.g., $\dot{a}<0$ would imply that older workers are more averse to strenuous effort than younger workers).

The firm's instantaneous profits are given by:

$$\Pi(t) = \left\{\frac{1}{b(t)} Q(h(t)) - w(t)\right\} e^{-rt} \tag{2}$$

$$Q'>0, \quad Q''<0$$

where Q is the production function, r is the market interest rate available to firms, and b(t) allows the worker's productivity to vary over the life cycle due to both the accumulation and depreciation of human capital.

Given this simple apparatus we can now define the optimal long term labor contract as follows. In a competitive labor market with identical firms, contracts offered by a firm must maximize the expected present value of utility of entering workers (otherwise the firm will attract no new workers) subject to three types of constraints. First, the present value of profits must equal zero. Second, rational expectations require that the equilibrium contract not be cheated on in any way by either party, that is, it is incentive-compatible because neither workers nor firms will agree to a (w(t), h(t)) path that they know cannot be realized because one party to the agreement will find it profitable to break that agreement later. Finally, while it is at least conceptually possible to have w(t) negative, we must impose a non-negativity constraint on h(t). Thus, when H is a hiring cost incurred at t=0 by the firm, the optimal contract satisfies:

$$\max_{w(t),\, h(t)} \int_{t=0}^{T} \left\{U(w(t)) - \frac{1}{a(t)} V(h(t))\right\} e^{-\delta t} dt \tag{3}$$

subject to

$$\int_{t=0}^{T} \left\{ \frac{1}{b(t)} Q(h(t)) - w(t)\right\} e^{-rt} dt - H \geq 0 \tag{4}$$

and both to $h \geq 0$ and the set of "cheating" constraints we develop in subsection 2 below. In equations (3) and (4), if disutility of effort, $\frac{1}{a(t)}$, rises with t and productivity, $\frac{1}{b(t)}$, declines after a certain age, contracts will tend to include a period near their end where the non-negativity constraint on h is binding. This will have h=0 but possibly w>0 and is of course the "retirement period" whose length is thus endogenized in the model.

2. The Effect of Malfeasance on Contracts

We shall consider the constraints imposed by Type 1 and Type 2 malfeasance in turn here.

(a) "Type 1" malfeasance

In order for a contract not to be vulnerable to quits by workers or dismissals by firms as means of avoiding contractual obligations, it must be true that, throughout the contract period, firms and workers both expect to do better during the remainder of the contract within the firm than they can do by leaving it. Formally this means the contract must satisfy:

to prevent worker malfeasance: $$\int_{\tau=t}^{T}\left\{U(w(\tau)) - \frac{1}{a(\tau)} V(h(\tau))\right\} e^{-\delta\tau}\, d\tau \geq PVW_t(H) - \gamma_1 - S \tag{5}$$

to prevent firm malfeasance: $$\int_{\tau=t}^{T}\left\{\frac{1}{b(\tau)} Q(h(\tau)) - w(\tau)\right\} e^{-r\tau} d\tau \geq -\varepsilon_1 \tag{6}$$

where S is a fixed job changing cost (such as search and travel expenses) borne directly and immediately by the worker.

In (5) and (6), $\gamma_1 \geq 0$ and $\varepsilon_1 \geq 0$ represent the long-run costs to workers and firms in terms of lost "reputation" of breaking contracts in this fashion. Thus ε_i would include a reduced ability by the firm to attract workers in the future, while γ_2 captures whatever "stigma" is attached to workers who change firms for this reason. Finally, $PVW_t(\cdot)$ is a kind of indirect utility function whose derivation for any given t is discussed more rigorously later. It gives the best present value of utility obtainable from a contract designed over the period from t to T, subject to the constraint that the present value of profits over that period equals $(\cdot)$, and subject to all the relevant malfeasance problems during that period. Thus $PVW_t(H)$ represents the best the worker can do if he leaves his present firm, re-incurs a hiring cost, and works for another firm that earns zero profits counting the initial hiring cost for the rest of his working life.

(b) "Type 2" malfeasance

Whether or not one party to a contract will attempt to provide less of "its" factor (or extort more of the "other's" factor) than contracted for depends on how the other party is expected to respond. The exact nature of this response is hard to predict because, at any point during the life of the contract the parties are essentially in a bilateral monopoly situation (this is guaranteed by the "Type 1" malfeasance conditions that no one wishes to withdraw, except in the degenerate case where the contract does not deviate from the spot market), so we shall proceed by analyzing two possible types of responses here. In one, the (potentially) injured party has been able to develop the perception of a credible threat that, should it detect malfeasance by the other party, it will immediately and unilaterally terminate the employment relationship, even if this action in itself is harmful to it. In the other the (potentially) injured party has not been able to make a believable threat and is expected to passively tolerate malfeasance up to the point where it is actually better off by leaving the firm. The first of these is modelled by Lazear (1979, 1981) in his version of the firm's behavior towards cheating workers to the author's knowledge the second and the distinction between the two have not been pointed out elsewhere. We deal with them in turn below.

When credible threats to withdraw from the contract when cheating by the other party is detected exist, the conditions needed to prevent "Type 2" malfeasance are:

$$\int_{\tau=t}^{T}\left\{U(w(\tau)) - \frac{1}{a(\tau)} V(h(\tau))\right\} e^{-\delta\tau}d\tau \geq PVW_t(H) + \theta - \gamma_2 - S \tag{7}$$

$$\int_{\tau=t}^{T} \left\{\frac{1}{b(\tau)} Q(h(\tau)) - w(\tau)\right\} e^{-r\tau}d\tau \geq \beta - \varepsilon_2. \tag{8}$$

which are identical to (5) and (6) except for θ, β, γ_2 and ε_2. The parameter θ represents the amount of worker cheating (measured in utility units) that workers are able to "get away with" before they are discovered and dismissed by the firm, while β represents the amount of cheating (in profit units) firms manage to fit in before workers discover them and terminate the employment relationship. The reputation parameters, γ_2 and ε_2, may differ from what they are in Type I cheating, since for example workers who are dismissed for shirking may find it harder to get another job than workers who quit.

When credible threats to terminate the contract do not exist, we may expect Type 2 malfeasance to be a more serious problem in contracts, since now an individual who cheates and is detected will not suffer dismissal --

unless he cheats by more than the amount needed to keep the other party indifferent between leaving the contract and not, which we assume is not his best policy. (If it were, the model would be just like when there are threats). This means that the conditions needed to prevent malfeasance are more stringent and of the form:

$$\int_{\tau=t}^{T} \{U(w(\tau)) - \frac{1}{a(\tau)} V\,(h(\tau))\}\, e^{-\delta\tau} d\tau \geq PVW_t(0) - \gamma_2, \tag{9}$$

$$\int_{\tau=t}^{T} \frac{1}{b(\tau)}\left\{Q\,(h(\tau)) - w(\tau)\right\} e^{-r\tau} d\tau \geq PV\Pi_t\{PVW_t(H) - S\} - \varepsilon_2 \tag{10}$$

which uses our indirect utility function, $PVW_t(\cdot)$, as well as the indirect profit function dual to it, $PV\Pi_t(\cdot)$. Thus $PVW_t(0)$ gives workers' gains to Type 2 malfeasance when there are no threats by firms, which is the maximum of ensuing utility within the firm subject to the constraint that firms will not want to "quit", i.e., ex post profits are at least zero. On the other hand, $PV\Pi_t[PVW_t(H)-S]$ gives firms' gains to cheating here, which are the most profits that can be earned from t to T that still leaves the worker indifferent between staying and moving to another firm that earns zero profits, where he receives a discounted utility of $PVW_t(H)$, but incurs a search cost of S in so doing.

3. The Competitive Context

The final step in outlining the structure of the model is to define more formally how the streams $PVW(\cdot)$ and $PV\Pi_t(\cdot)$ can be constructed, since it is these quantities that embody the constraints that a competitive labor market places on the structure of employment contracts. This definition of "best market alternatives" is possible only now that the description of all forms of cheating has been completed, and we can write out the definition of $PVW_t(H)$, for example, as:

$$PVW_t(H) = \max_{w(t),\, h(t)} \int_{\tau=t}^{T} \{U(w(\tau)) - \frac{1}{a(\tau)} V(h(\tau))\} e^{-\delta t} \tag{11}$$

subject to $\int_{\tau=t}^{T} \{\frac{1}{b(\tau)} V(h(\tau)) - w(\tau)\} e^{-rt} \geq H$, and to (5)-(10) and $h>0$.

Of course, $PVW_{t_1}(\cdot)$ will depend on $PVW_{t_2}(\cdot)$ for all t_2 strictly greater than t_1 because of the worker malfeasance constraints (5), and (7) or (9) that impinge on it: the way to derive the entire stream $PVW_t(\cdot)$ for $t>0$ needed to solve our problem is to work back recursively from time T, at which point all these constraints reduce to a condition on the spot wage versus the spot marginal product. There is no need to do this explicitly here of course; all we need for our purposes is to note two crucial properties of $PVW_t(\cdot)$: first, it can be properly treated as exogenous for all $t>0$ in the solution of our present optimal contract problem for $t=0$ to T, and second, $PVW_t(\cdot)$ is strictly decreasing in $(\cdot)$ because a higher minimum profit level reduces the present value of utility achievable from t to T in two ways -- both directly because of the higher profits over the whole period from t to T that must be allowed, and indirectly because higher profits at all subsequent t are also implied through the "cheating" constraints, thus restricting the scope of possible contracts. This means $PVW_t(H) < PVW_t(0)$, $\forall_t$.

The same kind of derivation may be done for $PV\Pi_t(\cdot)$ as for PVW_t but is not presented here for brevity. We simply note that, like PVW_t, $PV\Pi_t$ is decreasing in its argument, and exogenous for all $t>0$. Also note that duality implies $PV\Pi_t\{PVW_t(H)\} = H$ and thus by the monotonicity of PVW_t, $PV\Pi_t\{PVW_t(H)-S\} > H$.

IV. SOME BASIC IMPLICATIONS.

Some simple but important implications of our model can be discovered simply by examining constraints (5) - (10) without actually solving the optimal contract problem. Only two of the most important are examined here. They are: First, given equal reputation costs, type (2) malfeasance constraints always dominate type 1. Second, the conclusion that, in general, optimal contracts will tend to have wages rising faster than marginal products depends on a formulation in which worker cheating is Type 2 but firm cheating is Type 1, and disappears when firms are allowed to cheat in the smae manner as workers. In making this last point, we assume Type 2 cheating is such that threats are available to both parties to the contract. This is simply for brevity of exposition since it is easily shown that allowing for the possibility that cheated partners react passively either strengthens or leaves unaltered that conclusion.[1]

1. Type 1 versus Type 2 malfeasance: to withdraw or to exploit?

Whether it is Type (1) or Type (2) malfeasance that effectively constrains the structure of contracts depends of course on which of these is more attractive to the party which cheats. A quick look at (5) - (10) reveals that, in our simple world here, Type (2) malfeasance, or "*ex post* exploitation" will always provide the binding constraint if the reputation costs of the two forms of malfeasance are equal (i.e., $\gamma_1=\gamma_2$ and $\varepsilon_1=\varepsilon_2$). This is because, for workers, $PVW_t(0) > PVW_t(H)$, $\forall t$, so the right-hand side of both (7) and (9) always exceeds that of (5), meaning that when $\gamma_1=\gamma_2$ constraint (5) is dominated by whichever of the other two is operative. A similar argument can be made for firms. The intuitive reason why

1. See a more complete, earlier version of this paper, available on request.

exploitation is preferred over simple withdrawal is clear: you can only get your own best market alternative by withdrawing; you can get more by exploiting the other party, even if you are eventually "dismissed" as a result.

While it is possible, of course, for $\gamma_2>\gamma_1$ or $\varepsilon_2>\varepsilon_1$ by enough to make Type 1 malfeasance the most likely -- say fired workers find it much harder to get new jobs than quitters and firms which have exploited workers have a harder time finding new workers than others who simply fired some other workers in, perhaps, a business slump -- we shall focus for the remainder of this paper on Type 2 cheating only, since it is here that the <u>new</u> implications of letting firms "exploit" just like workers "shirk" are brought out.

2. <u>Will wages tend to rise faster than marginal products?</u>

To see how the conclusions that wages will tend to rise faster than marginal products in optimal contracts depends on a model where worker cheating is Type 2 but firms may cheat only by dismissing workers (Type 1), imagine the binding constraints are (6) and (7) and examine what they say about required contract structure. Equation (6) states that, for contracts to prevent firm malfeasance, all we need is for <u>ex post</u> rents to the firm (the difference between the left hand side and zero) to not be too negative. Thus, while there is a need for wages not to rise too fast relative to productivity, there is <u>nothing to be gained in this world by making ex post rents to the firm positive</u> -- i.e., essentially by setting up a situation where wages rise more slowly than productivity. Since (7) stipulates that <u>ex post</u> rents to workers (its left-hand side minus PVW(H)-S) must exceed $\theta-\gamma_2$, which may be positive, there is however a reason to make wages rise more quickly than productivity, especially if workers have little or no reputation, and optimal contracts will tend to lie in this direction. This is Lazear's (1979, 1981) result; the intuition is that contracts must hold back wages by more than the amount workers can extort from firms by cheating, but they do not need to hold back profits because all firms can do is dismiss workers, which only requires that *ex post* profits *not be too negative*.

When firms are allowed to cheat just like workers do and (8) replaces (6) as the binding constraint, the above argument no longer holds of course because both parties can now gain positive amounts by cheating, so *both ex post* rents to workers and profits must be positive in equilibrium contracts (which is possible in a competitive market with zero *ex ante* rents and profits because of hiring costs, search costs, and job-specific human and physical capital), and there is no clear reason, *ex ante*, to expect wages to rise faster than marginal products or *vice versa*. While this result might still be the best policy, that of course depends on such factors as reputation, discount rates, and so on, as we shall examine later.

V. THE OPTIMAL CONTRACT: GENERAL SOLUTION.

The optimal contract problem as represented so far is not amenable to standard dynamic optimization techniques, but a simple transformation of variables suffices to turn it into a fairly conventional calculus of variations problem with constraints. To do so, we define four state variables -- μ, ν, q and ω -- equal to the remaining present values at time t of utility of income, disutility of effort, output, and wages, respectively. Thus we have:

$$\begin{aligned}
\mu(t) &= \int_{\tau=t}^{T} U(\tau)\, e^{-\delta\tau} d\tau \\
\nu(t) &= \int_{\tau=t}^{T} \frac{1}{a(\tau)} V(\tau)\, e^{-\delta\tau} d\tau \\
q(t) &= \int_{\tau=t}^{T} \frac{1}{b(\tau)} Q(\tau)\, e^{-r\tau} d\tau \\
\omega(t) &= \int_{\tau=t}^{T} w(\tau)\, e^{-r\tau} d\tau
\end{aligned} \qquad (12)$$

This allows us to write the optimal contract problem as the simple calculus of variations problem:

$$\max_{\substack{\mu(t),\ \nu(t),\\ q(t),\ \omega(t)}} \int_{t=0}^{T} (\dot{\nu} - \dot{\mu})dt \tag{13}$$

subject to a zero-profit constraint:

$$\int_{t=0}^{T} \{\dot{\omega} - \dot{q} - H\}dt = 0 \tag{14}$$

subject also to all the "malfeasance constraints" which we summarize by the two conditions:

worker cheating $\mu - \nu \geqq K_w$ (15)

firm cheating $q - \omega \geqq K_f$ (16)

where K_w and K_f vary both according to which of (5) - (10) is binding and according to the levels of the parameters within the binding constraint, and finally subject to the functional relationships between $\dot{\omega}$ and $\dot{\mu}$ and $\dot{q}$ and $\dot{\nu}$:

$$U(-\dot{\omega}\, e^{rt}) = -\mu e^{\delta t} \tag{17}$$

$$Q(V^{-1}\{-\dot{\nu}\, a(t)e^{\delta t}\}) = -\dot{q}b(t)e^{rt} \tag{18}$$

as well as h>0.

To further simplify notation, we shall denote $Z(\cdot) = Q\{V^{-1}(\cdot)\}$ and note that it is a strongly concave function of the utility of effort, V, because both A and V^{-1} are concave, and that Z'' becomes more negative as Q'' falls and V'' rises. Denoting λ_w and λ_f as the multipliers for (15) and (16) respectively and γ_u and γ_q as the multipliers for (17) and (18) respectively, the first-order conditions for a constrained maximum of (13) can be written:

$$\lambda_w - \dot{\gamma}_u e^{\delta t} - \gamma_u \delta e^{\delta t} = 0 \tag{19}$$

$$-\lambda_w + \dot{\gamma}_q \, a e^{\delta t} Z'(V) + \gamma_q\{(\dot{a}e^{\delta t} + a\delta e^{\delta t})Z'(V) + ae^{\delta t}Z''(V)\dot{V}\} = 0 \tag{20}$$

$$\lambda_f - \dot{\gamma}_q b e^{rt} - \gamma_q(\dot{b}e^{rt} + bre^{rt}) = 0 \tag{21}$$

$$-\lambda_f + \dot{\gamma}_u\{re^{rt}U'(w) + e^{rt}U''(w)\dot{w}\} = 0 \tag{22}$$

$$\lambda_w\{\mu - \nu - K_w\} = 0 \tag{23}$$

$$\lambda_f\{q - \omega - K_f\} = 0 \tag{24}$$

along with (17) and (18). Equations (19) through (24) are just the Euler equations for μ, γ, q, and ω respectively. The easiest way to characterize the optimal contract $(w(t), h(t))$ from these conditions is to solve them for $w(\ )$ and $V(t)$ and then note that h is just a monotonic function of V from $V = V(h)$. Solving (19) and (21) for $\dot{\gamma}_q$ and $\dot{\gamma}_u$ and substituting into (20) and (22) we then get:

$$\dot{w} = \left\{\frac{-1}{\gamma_\mu e^{rt}U''(w)}\right\} \cdot \left\{-\lambda_f + \frac{\lambda_w - \gamma_\mu \delta e^{\delta t}}{e^{\delta t}} e^{rt}\, U'(w) + \gamma_\mu r e^{rt} U'(w)\right\} \tag{25}$$

$$\dot{V} = \left\{\frac{-1}{a\gamma_q e^{\delta t} Z''(V)}\right\} \cdot \left\{-\lambda_w + \frac{\lambda f - \gamma q(\dot{b} + br)e^{rt}}{be^{rt}}\, ae^{\delta t}Z'(V) + \gamma_q(\dot{a} + a\delta)e^{\delta t}Z'(V)\right\} \tag{26}$$

which are just first-order differential equations for $\dot{w}$ and $\dot{V}$.

Examination of (25) and (26) reveals the following:

First, when neither firm nor worker-cheating constraints are binding ($\lambda_f = \lambda_w = 0$), the optimum contract is given by:

$$\dot{w} = \left(-\frac{U'}{U''}\right)\cdot(r-\delta) \tag{27}$$

$$\dot{V} = \left(-\frac{Z'}{Z''}\right)\cdot\left(\delta - r + \frac{\dot{a}}{a} - \frac{\dot{b}}{b}\right) \tag{28}$$

Equation (28) can be written directly in terms of h using the inverse function rule and the definition of Z as :

$$\dot{h} = \frac{1}{\frac{V''}{V'} - \frac{Q''}{Q'}}\left(\delta - r + \frac{\dot{a}}{a} - \frac{\dot{b}}{b}\right) \tag{29}$$

We call this the first-best contract since it represents the best that can be done if malfeasance is not a problem in designing contracts. In this contract the wage profile is totally independent of the productivity profile, is rising if firms discount the future more than workers and falling otherwise, and the absolute value of its slope diminishes with the degree of absolute risk aversion of workers. The first-best effort profile is more likely to be rising if workers discount the future more than firms $(\delta - r) > 0$, will tend to be rising in periods when workers are becoming more productive $(\frac{\dot{b}}{b} < 0)$ and will tend to be falling when workers value leisure increasingly $(\frac{\dot{a}}{a} < 0)$. The variation of effort over the life cycle diminishes as Q becomes more concave and V becomes more convex (i.e., as it becomes more desirable both from a productive and utility standpoint to smooth labor input over time). Note that the effort profile is also independent of the income profile.

Second, if either the worker-cheating constraint (15) or the firm cheating constraint (16) is binding (both are so at once only in the degenerate, spot market case), the actual second-best labor contract deviates from the first-best one. When the desired contract is such that there are periods and wage levels where workers' deferred utility within the firm is not high enough to prevent worker cheating (i.e., $\lambda_w > 0$), $\dot{w}$ tends to be higher there and h lower, which acts to increase worker's deferred utility. Similarly, whenever firm cheating is a problem in the desired contract ($\lambda_f > 0$), wages tend to rise less (or fall more) and effort tends to rise more (or fall less). In short, while the first-best contract is able to both allocate income and effort over the life cycle in an efficient manner, both independently of each other, the actual second-best contract is constrained in its ability to do so. Income profiles are constrained by productivity profiles in order to prevent cheating because the firm can never "owe" the worker too much and vice versa, so income profiles are likely to track productivity profiles much more closely than they would otherwise. The allocation of effort over the life cycle is also influenced by cheating considerations instead of depending solely on when effort is most valued by the firm and least disdained by the worker.

VI. AN APPLICATION: UNIONISM.

To see more clearly the kinds of implications the above model has for the shapes of compensation and effort profiles, it helps to postulate some reasonable functional forms, solve out an example, and apply it to a concrete empirical issue. We shall look at the probable effects of unionism on employment contracting here in three stages: First, the way unionism might be expected to affect the parameters of the problem, specifically K_w and K_f, is discussed. Second, the assumptions of our particular example are set out, and the structure of first-best contracts is derived. Finally the probable structure of actual, second-best, union and nonunion contracts is more informally characterized and these predictions are compared with the empirical evidence.

1. Effect of Unionism on Parameters

Unionism can concievably have many different impacts on the components of the parameters K_w and K_f in our model. These include: unions' effects on whether the constraints embodying threats or the others which assume passive responses to malfeasance are operative, their effects on overall rents to being in the firm, and finally some probably less-important factors that have been mentioned elsewhere in the literature. We discuss these in turn below.

(a) are threats operative?

We propose as a working hypothesis that, while any firm of a reasonable size is likely to have access to threats over malfeasant workers, it is only in unions where workers can enforce credible threats over firms. The main reason for this is simply that the *ex post* bargaining game that occurs because of the bilateral monopoly situation within contracts is played many times by a typical firm but only once by a typical worker: To prevent worker shirking, firms can make "examples" of malfeasant workers through disciplinary dismissals and still gain in the long run because they acquire a reputation for doing so among their own labor force. To prevent workers making all-or-nothing wage demands, firms can categorically refuse to make individual wage bargains, thus for example attaching wages to jobs, not workers, activities which Williamson *et al*. (1975) have labelled "internal labor market" policies. Individual workers have access to no such methods, since firms know that actually exercising a threat does not make sense if you expect to play a game only once and therefore cannot recoup the losses from doing so in terms of better contract performance later. On the other hand, unions, because of their large membership and longer lifespan than individual workers may well be able to establish a reputation for "punishing" malfeasance by firms, either by a withdrawal of labor services (i.e., a strike) or probably more commonly through less drastic forms of pressure.

Loosely, then, we can represent one probable effect of unions as the replacement of constraint (10), which implies workers' only response to firm malfeasance is to quit if they can do better elsewhere, by constraint (8) which implies they withdraw their services whenever any firm malfeasance is detected, implying a lower K_f in unions.

(b) *rents and union wage effects*

It is fairly widely accepted that unions have a significant, positive overall total wage effect. This means that, *cet. par*, union members probably have more to lose by losing their jobs (PVW inside the firm -- PVW outside is greater for any given contract structure), which is the equivalent of a lower K_w.

(c) *other effects*

Some other comments about unions' effects on contracts have been made in the literature; these tend on the whole to fit in rather nicely with the previous hypothesis. For one, Freeman (1981) has suggested that unions help to realize important scale economies in the provision of information regarding employer compliance with complex contract provisions. This only strengthens our notion that unions lower K_f due to an ability to discipline the employer. Also, Lazear (1981) has suggested that unions, again unlike individual workers but like firms, may acquire with their longer lifespans a reputation for delivering labor services of a certain quality. If unions do tend to screen and police their own members in this way, it reinforces the conclusion that union overall wage effects are likely to lower K_w. Finally a third factor has been commented on that does <u>not</u> fit with our proposition that unions lower both K_w and K_f -- this is the notion that unions may interfere with firms' ability to make <u>genuine</u> disciplinary dismissals, which is a factor that will tend to raise K_w in unions. We shall not include this in our maintained hypothesis here, noting that it would have to outweigh the two other negative effects on K_w -- especially the greater loss should a dismissal occur -- to make a difference, and if this is actually true, it will become apparent when the present example's predictions are compared with the evidence.

2. <u>Assumptions of the Example and Derivation of First-Best Contracts</u>

To make it easy to characterize the shape of first-best contracts, let us make the convenient assumption that utility, disutility-of-effort, and production functions all have the simple constant absolute risk aversion forms, $f(x) = a^x$. This means $\frac{U'}{U''}$, $\frac{V'}{V''}$, and $\frac{Q'}{Q''}$ are constants, equations (27) and (29) no longer depend on w and h respectively but only on t, and the optimal wage profile is linear. To further simplify matters, imagine $\delta = r$, so that first-best wage contracts are completely flat ($\dot{w}=0$), which is of course the standard result in unconstrained income-allocation problems of this nature. Next, although we might believe workers' value of leisure on the job increases as they get older, assume for simplicity that $a(\tau)$ is constant even though letting it rise strengthens the conclusions regarding wages and effort of older workers we derive below. Finally, let us follow the predictions of human capital theory (e.g., Ben-Porath (1967)) which state that optimal investment in human capital over a finite lifetime with depreciation is likely to generate a productivity profile, $1/b(t)$, that at first rises and then falls near the end of working life. Assuming the optimal h always exceeds zero and thus ignoring retirement issues, the first-best effort profile is now given by:

$$\dot{h} = -\frac{1}{R_q - R_v} \cdot \frac{\dot{b}}{b} \tag{30}$$

where R_q and R_v are the coefficients of "absolute risk aversion" of the V and Q functions respectively (R_v is negative because V is convex.) Since b(t) by assumption is at first decreasing and then increasing, this means the optimal effort profile will have the shape shown in Figure (1)(a). The worker is asked to deliver more effort in the prime of working life when most productive, less when still learning, and also less when older.

3. Second-Best Contracts in Union and Non-Union Firms

The easiest way to characterize union versus non-union contracts is first to analyze how second-best contracts deviate from first-best ones, and then to note that this deviation will be less in union than non-union firms since we have proposed that unions "loosen" both the constraints ((15) and (16)) that make contracts deviate from their first-best level. We do this in turn below and conclude with a brief discussion of the empirical evidence on the question.

Figure 1(c) shows how second-best contracts are likely to deviate from first-best contracts by indicating where the first-best contract violates the malfeasance constraints. It shows roughly what the first-best path of remaining present value of utility will look like (derived from Figure 1(a) by first depicting the instantaneous utility stream it implies--Figure 1(b)--and then taking present values) and indicates it is likely to violate worker malfeasance constraints (directly from equation (7) or (9)) near the beginning of the contract, and firm malfeasance constraints (derived by putting the minimum PV Π levels in (8) or (10) into the indirect utility function $PVW_t(\cdot)$) near the end of the contract.

What does this mean? Near the end of the contract, workers' productivity is low and falling, and it is optimal to utilize them less without lowering their compensation. But this means the remaining profits from employing the worker are low and probably negative, so firms will cheat. Second-best contracts will compensate for this by lowering $\dot{w}$ and raising $\dot{h}$ from their first-best levels, as is indicated when λ_f is positive in (25) and (26). In other words, second-best contracts underpay and overwork older workers in order to reduce firms' incentives to break those contracts. Near the start of the contract, workers should have already received an "advance" on their future productivity in first-best contracts and now face a period where their productivity and effort levels rise but their compensation remains flat. But this means *ex post* utility is relatively low, so workers will cheat. Second-best contracts compensate for this by raising $\dot{w}$ and lowering $\dot{h}$ of younger workers, as a positive λ_w in (25) and (26) indicates. Thus, second-best contracts give young workers less of an "advance" on their future productivity (in other jargon, pay for a smaller

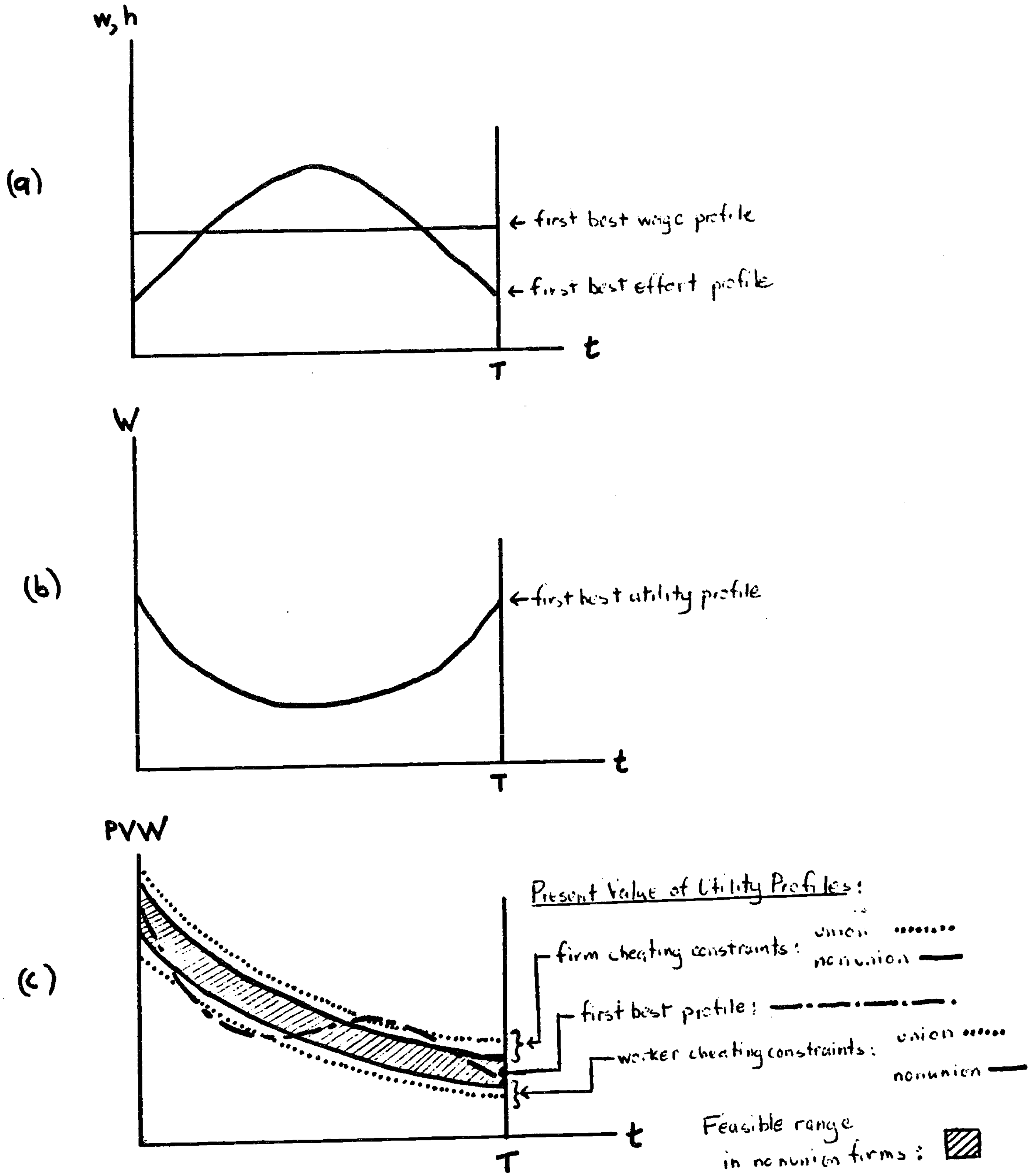
w, h
(a)
← first best wage profile
← first best effort profile
t
T
W
(b)
← first best utility profile
t
T
PVW
(c)
Present Value of Utility Profiles:
firm cheating constraints: union
nonunion
first best profile:
worker cheating constraints: union
nonunion
t
T
Feasible range in nonunion firms:

Figure 1

fraction of workers' training) in order to reduce workers' incentives to break those contracts.

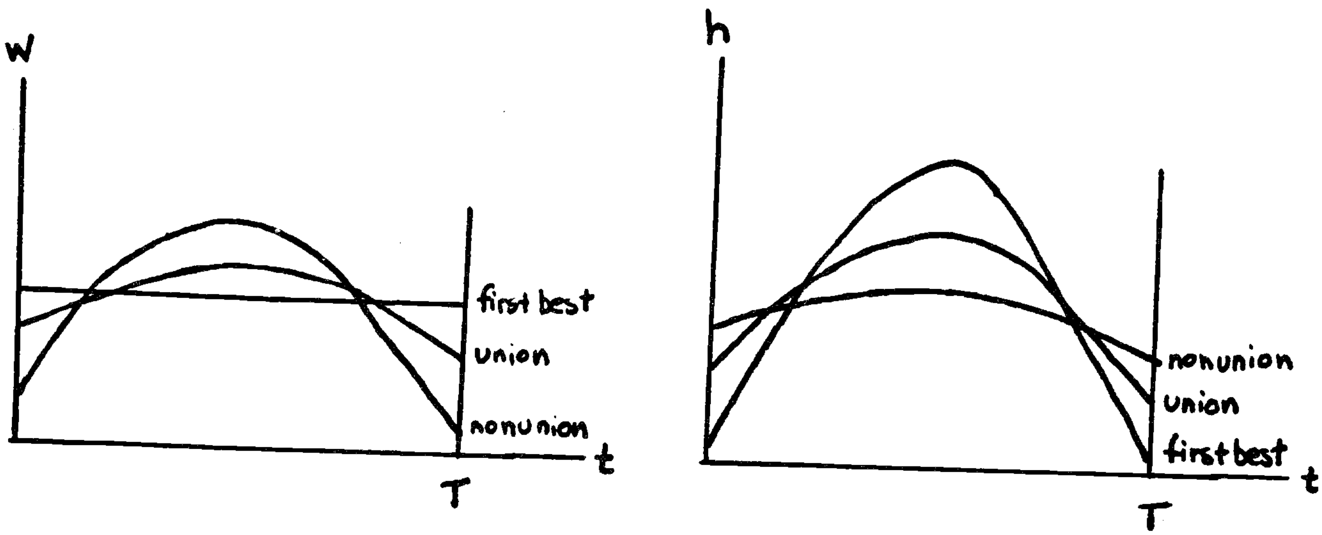

Figure 2

Predicted union-non-union differences from the model follow naturally by noting how Figure 1(c) shows λ_f and λ_w are likely to be positive over longer periods in non-union that in union firms, so profiles like those shown in Figure (2) emerge: Union compensation profiles are less peaked, and union effort profiles are more peaked than their non-union counterparts. In both cases these are closer to the first-best profiles since unions can mitigate both the firm-cheating problem--allowing compensation to be sustained more and effort to drop off more for older workers--and the worker malfeasance problem--allowing the youngest workers to be given larger "advances" of compensation and the most productive (middle-aged) workers in the firm to be utilized more intensely.

Empirical evidence on union versus non-union compensation and effort profiles is not extensive but broadly consistent with the example's predictions. The fact that union wage profiles appear to be less peaked than non-union profiles has been documented in Johnson (1981) and Johnson and Youmans (1971), who find a U-shaped union wage effect by age. Indeed, this paper is apparently the first to provide a consistent, choice-theoretic explanation of that phenomenon. The finding by Freeman (1981) that unionized firms tend to spend more on pensions fits in with this trend. Regarding effort profiles, one might interpret certain work rules as evidence of unions' being able to maintain lower effort levels of older workers, which are preferred _ex ante_ by workers writing contracts but unachievable in non-union firms, although more concrete evidence such as an earlier retirement age for union members would be better evidence of this trend. As to whether the effort profile is more positively sloped for younger union workers, no evidence suggests itself, but it is also true that this prediction of the model is an artifact of the fact that we have not built in growing aversion to effort with age.

In short, the predictions of this simple example fit in quite well with the known empirical facts about unions (and could be made to fit even better, for example by including a secular earning growth trend due to $r > \delta$ and by letting disutility of effort rise with age) giving some support to the proposition that unions may have positive economic functions by acting as workers' equivalents to disciplinary dismissals policies in enforcing contracts, and to the general model of employment contracting developed here as a whole. Indeed, the analysis has some interesting corollaries--for example, that unions in what is probably their main role of "policers" of firms are more likely to be useful in situations where older workers are more vulnerable to malfeasance by firms, which might help explain (a) the relative concentration of unionism in manual, blue collar occupations where productivity probably depreciates more quickly as the worker approaches retirement, and (b) the decline in unionization during the postwar period, when a number of government programs that both protected workers from, and made them less vulnerable to, firm malfeasance multiplied and expanded (consider especially ERISA, OSHA, and UI here). Has the government taken over the role of unions?

References

Azariadis, C., "Implicit Contracts and Underemployment Equilibria," Journal of Political Economy (December 1975), pp. 1183-1202.

Baily, M.N., "Wages and Employment Under Uncertain Deamnd," Review of Economic Studies (January 1974), pp. 37-50.

Baumol, W.J., "Contestable Markets: An Uprising in the Theory of Industry Structure," American Economic Review (March 1982), pp. 1-15.

Ben Porath, Y., "The Production of Human Capital and the Life Cycle of Earnings," Journal of Political Economy, No. 4, (1967), pp. 352-365.

Eaton, J. and H.S. Rosen, "Pensions, Agency and the Structure of Executive Compensation," mimeo, June 1981.

Freeman, R., "The Effects of Trade Unionism on Fringe Benefits," Industrial and Labor Relations Review (July 1981).

Grossman, H.I., "Risk Shifting, Layoffs, and Seniority," Journal of Monetary Economics, (November 1978), pp. 661-686.

Hall, R., and E. Lazear, "The Excess Sensitivity of Quits and Layoffs to Demand," NBER Working Paper No. 864, (February 1982).

Harris, M., and B. Holmstrom, "Ability, Performance and Wage Dynamics," mimeo, 1981.

Hashimoto, M and B.T. Yu, "Specific Capital, Employment Contracts, and Wage Rigidity," Bell Journal of Economics (Autumn 1980), pp. 536-549.

Holmstrom, B., "Equilibrium Long-Term Labor Contracts," Northwestern University, Kellogg Graduate School of Management, Discussion Paper No. 414, (January 1980).

Johnson, G. and K. Youmans, "Union Relative Wage Effects by Age and Education," Industrial and Labor Relations Review (January 1971), pp. 171-179.

Johnson, G., "Changes Over Time in the Union-Nonunion Wage Differential in the United States," mimeo, 1981.

Klein, B. and K. B. Leffler, "The Role of Market Forces in Assuring Contractual Performance," Journal of Political Economy (August 1981), pp. 615-641.

Lazear, E., "Why is There Mandatory Retirement," Journal of Political Economy (December 1979), pp. 1261-1284.

----------, "Agency, Earnings Profiles, Productivity and Hours Restrictions," American Economic Review (September 1981).

Lazear, E., and H.S. Rosen, "Rank-Order Tournaments as Optimum Labor Contracts," Journal of Political Economy (October 1981).

Medoff, J. and K. Abraham, "Involuntary Terminations Under Explicit and Implicit Employment Contracts," mimeo, 1981.

Williamson, O. et al., "Understanding the Employment Relation," Bell Journal of Economics (Spring 1975), pp. 250-278.