Volume Title: Pensions, Labor, and Individual Choice

Volume Author/Editor: David A. Wise, ed.

Volume Publisher: University of Chicago Press

Volume ISBN: 0-226-90293-5

Volume URL: http://www.nber.org/books/wise85-1

Publication Date: 1985

Chapter Title: Insurance Aspects of Pensions

Chapter Author: Peter A. Diamond, James Mirrlees

Chapter URL: http://www.nber.org/chapters/c7138

Chapter pages in book: (p. 317 - 356)

11 Insurance Aspects of Pensions

Peter A. Diamond
James A. Mirrlees

11.1 Deferred Wage Payments

If the State does not provide an adequate system of pensions, it is to be expected that employers will provide their workers with pensions, and some do. In a world of full information and perfect markets, private pensions could be individual contracts by workers with independent insurance companies. We shall not comment on all the reasons why that does not happen. Instead, we consider the forms of private pensions that are likely to arise when the employer provides the pension. We do this as an alternative to analyzing optimal pension policy by a benign State and comparing it with features of what currently exists. It should be possible to identify ways of regulating or supplementing private pensions that are likely to increase welfare.

The first task is to capture the important features of company pensions in a model that is easy to think about. To identify some of these features, we begin by asking why pensions exist, which is to say, why some part of what might have been paid as an immediate wage may be converted into future pension rights. (This should throw some light on the reasons the employer does it for the worker, rather than leaving it to the worker's independent arrangements.) Then we ask what might be expected to limit the extent of deferred payments. We are led to formulate a simple three-period model, which is used in sections 11.2–11.4 to see how the limitations referred to affect pensions in the model. In section 11.5 we ask how well private pensions in the model would provide for early retirement. The remaining sections discuss informally some issues the simple model does

Peter A. Diamond is professor of economics at Massachusetts Institute of Technology. James A. Mirrlees is professor of economics at Oxford University.

not illuminate: the consequences of inflation, the form of pension plans, and the effects of repeated wage and pension negotiations.

Why, then, might some part of the wage be deferred? In the first place, it might be done to provide saving. In the past there have often been tax advantages to saving channeled through an employer's pension plan. In any case, a higher return on capital might be available to the larger investments handled by the employer, particularly since marketing costs are low in this context. A further possibility is that deductions from the wage provide a particularly effective commitment to save that could be attractive to workers and their representatives who are apprehensive of future weakness of will. We do not propose to model any of these considerations here, since their bearing on pensions is easy to understand and their existence is unlikely to affect the desirable form of a pension plan, given that some plan is desirable.

There is a second class of reasons for deferring part of the wage, namely, to provide various kinds of insurance. We assume workers to be rational, an assumption we shall make throughout the paper; thus they would like to have insurance against many work-related contingencies. The first and most obvious is disability. More generally, there is substantial uncertainty for any worker about his future productivity. We want to distinguish between uncertainty about his productivity generally and uncertainty about his productivity in the current employment relative to his productivity elsewhere in the economy. This distinction will later turn out to be quite significant. These uncertainties relate to what might be directly observed in due course, wages to be earned in the future. There are also important uncertainties about the desirability of changing employment for reasons other than wage differences, because of, say, changing relationships within the firm, with people elsewhere. Indeed there is always uncertainty about the relative intertemporal value of consumption, quite apart from desires to change employment. Being related, often, to events that are hard to verify, these uncertainties cannot be directly the object of insurance.

A worker would therefore like to put some part of his current full wage into providing income in contingencies when it will be more important to have. Provision of a pension on final retirement may be a convenient way of getting his insurance, if he can expect to make a small contribution to the retirement and disability pensions when he has a lower full wage. It is not impossible that, at some stage in his working life, he would want to reduce his guaranteed retirement pension and consume more than his full wage currently: depending on the wage history, he might want a negative pension or to borrow against pension rights.

This last consideration suggests that the level of pension contributions should vary, perhaps greatly, with the worker's individual experience. Why then should we observe that workers in employment generally have their pension plan arranged by the employer rather than by an independent insurance company? The obvious answer, and probably the right one, is that it pays insurance companies to arrange insurance with groups of workers rather than with individual workers, both because it saves marketing costs (which are remarkably high for individual insurance) and because it limits the impact of adverse selection. It is hard to see how these advantages can be obtained other than by pension plans universal within a company, an occupation, or large categories within a nation.

Since the providers of pension plans have done so much to eliminate the adverse selection problem, we shall take advantage of that to confine our attention (other than some remarks in section 11.5) to the two other issues that reduce the extent of insurance: unobservability of events that ideal insurance contracts would relate to, and the inability of workers to commit themselves to future payments. The reason that leads a worker to resign or retire may be unobservable: ideal insurance would pay more to a worker who becomes unable to work than to a worker who could work but chooses not to, and it would similarly avoid the problem of inducing workers to leave for pleasanter jobs elsewhere by making pensions conditional on the reasons for mobility. It turns out that the problem of observability interacts with the problem of commitment. The issue of commitment arises with all future contracts and is controlled by law and by reputation. Even when the law allows enforcement of promises to pay, it may be costly to ensure compliance: the cost in reputation of breaking a promise is therefore always important. There are also other observable costs, such as social disapproval—more important within some groups and classes of commitments than in others—which we shall ignore.

Considerations of reputation suggest that believable and effective commitments are easier for firms to make than for workers. We discuss this issue further in section 11.8 at the end of the paper. It might be possible for the State to provide cheaply an effective mechanism to ensure compliance with commitments made by individual workers, for example by introducing voluntary commitments to taxes or social insurance contributions. But, on the whole, the appropriate assumption for a realistic model is that workers are unable to commit themselves to make future payments to employers, whereas employers are able to commit themselves to make future payments to workers. We shall assume, for example, that a worker cannot enter into arrangements that ensure he does not benefit from accepting a higher wage offer from a competing employer, although he might have preferred to use the proceeds from such an agreeable contingency in less satisfactory states of nature.

These considerations suggest that we model the labor market as follows. There are three periods for the worker. In the first period, a wage contract is agreed between the worker and the firm employing him. That contract specifies the payment to be made in the first period and payments to be made by that firm in subsequent periods, contingent on various possibilities, such as whether the worker remains with the initial employer and whether he works in the second period. The payments specified refer to both of the remaining periods. In the third period, the worker is retired. Our economy is competitive: workers and firms compete for one another. But workers may have different productivities in different firms—this is an issue in the second period. In the first period, we shall suppose, neither worker nor firm knows what the worker's productivity or alternative wage opportunities will be in the second period. For the second period, we shall consider explicitly a number of the possible patterns of knowledge and observation: the worker may or may not know his marginal productivity to the firm he was with in the first period, and the firm may or may not be able to rely on observation of the wages available to the worker in alternative employments. An interesting hybrid case is where the wage in alternative employment is observable for determining the level of a transferable pension but not for determining the wage should the worker remain with his first-period employer.

Whatever our assumption about observability, we find it helpful to distinguish three mobility possibilities. The first is that, in equilibrium, workers never in fact change employment because their productivity is the same whichever firm they work for, although they might have moved, and equilibriuim contracts take account of that possibility. This situation has been studied by Harris and Holmstrom (1982) in a model with many time periods: we call it the case of *mobility threats.* Their model is the starting place of our analysis. The second case is that of *exogenous mobility*, where, for reasons not subject to objective verification, workers may wish to move to other employments (or to none). In this case, employment contracts are drawn up with full awareness that insurance for such an eventuality might encourage mobility also when the unobservable reasons do not apply. The remaining case is called *endogenous mobility,* where workers may find that better wages are available from other firms and the first employer may, because they have greater productivity elsewhere, be willing to allow them to move. In this case, it may or may not be possible for the first employer to observe the alternative wage offers. If it is not possible, there is still this distinction between endogenous and exogenous mobility, that a rational worker wants to insure against the latter happening and against the former not happening. Necessary early retirement because of disability is the extreme case of exogenous mobility.

11.2 Mobility Threats

We first set up the model and show it working in the case of mobility threats. It is our policy to simplify wherever possible in order to make results more vivid. This leads us to specify the same utility of consumption in each period, *u(c),* with the disutility of labor, of alternative employers, or of disability being subtracted from the utility of consumption as appropriate. We assume that u is increasing, concave, and differentiable with u' tending to $+\infty$ and zero as consumption goes to zero and $+\infty$. We also assume a zero real interest rate. In a world ideally well informed, consumption would then be the same in all periods and all states of nature. We focus on deviations from this ideal constancy.

The marginal productivity of labor with the first employer is denoted by m_1 and m_2 in the two periods when there is work. When we come to consider alternative employers, we shall denote marginal productivity there by m_2'. Wages actually paid are w_1, w_2, and w_2', respectively. Workers have no access to a capital market (an extreme assumption made because we are emphasizing insurance aspects, and the presence of savings complicates matters in some ways, without significantly changing anything that matters.[1]) Firms can freely borrow and lend at a zero real interest rate (an obvious simplification so long as we are not going to look at the effect of interest rate changes). Since no problems of observability or commitment attach to the distribution of payments to workers between the second and third periods, with workers all receiving payments in the third period, and all known to be retired in that period, we may as well aggregate the total payment received in these two periods. Whatever it is, it will in fact be divided equally between the two periods.[2] We therefore write b, b' for the contribution to pension arising from first-period employment, indicating that it may be different if the worker moves to a new occupation in the second period. Total worker income, from wage and pension, in the second and third periods is $w_2 + b$ or $w_2' + b'$, depending on whether he remains with his first employer or not. We write $W = w_2 + b$, $W' = w_2' + b'$, $U(W) = 2u(W/2)$.

Firms are taken to be risk neutral. Then in equilibrium, expected payments to a worker, including deferred payments, equal his expected marginal product. In competing for workers, firms must devise contracts that maximize the expected utility of workers, always having regard to mobility possibilities.

In the case of mobility threats, $m_2 = m_2'$. If a worker moved in the second period, he could obtain from his new employer income equal to m_2'. If his initial labor contract made that worthwhile, it must pay the first employer to change the contract without disadvantage to the worker. Thus the equilibrium contract is given by

$$\max_{w_1, W(\cdot)} \quad u(w_1) + EU[W(m_2)] \tag{1}$$

subject to
$$w_1 - m_1 + E[W(m_2) - m_2]. = 0$$
$$W(m_2) \geq m_2$$

Notice that m_2 is a random variable, to which the expectation operator applies.

If m_2 is such that $W(m_2) > m_2$, the second constraint does not apply and marginal utilities are equated: $W = 2w_1$. Otherwise the constraint determines the wage. Therefore the solution is

$$(2) \qquad W(m_2) = 2w_1, \quad m_2 \leq 2w_1,$$
$$= m_2, \quad m_2 \geq 2w_1,$$

with the first constraint in (1) determining the level of w_1. This is just the Harris-Holmstrom result, modified by the presence of a retirement period. The Harris-Holmstrom result was that wages do not fall but may rise if marginal product increases enough. If we had a two-period model, with $U(W) = u(W)$, we would get the result that the second-period wage is higher than the first-period wage if the second-period marginal product is higher than the first-period wage. The existence of a retirement period allows the first employer to pay a pension only if the worker has remained with him. Therefore the worker's current income can remain constant with greater probability, increasing for the second and third periods only if $m_2 > 2w_1$.

Using nontransferable pension rights becomes less attractive when mobility may be desirable.

11.3 Exogenous Mobility

We introduce exogenous mobility by specifying a positive probability p of leaving the first employer by "necessity." We maintain the assumption that a worker who is not under the necessity of moving would have the same marginal product in alternative employments as with the first firm. It is then a constraint on the contract that such a worker should not be induced to move. Denoting the marginal product (elsewhere) of a worker who has to move by m', we see that the equilibrium contract is defined by

$$(3) \qquad \max_{w_1, W(\cdot), b'} \quad u(w_1) + (1 - p)EU[W(m_2)] + pEU(m' + b')$$

subject to $w_1 - m_1 + (1-p)E[W(m_2) - m_2] + pb' = 0$

$$W(m_2 \geq m_2 + b', b' \geq 0.$$

Notice that we allow for a transferable pension component, b', which however is independent of m': that is, the new wage is supposed to be unobservable. The third constraint, that $b' \geq 0$, expresses the inability of workers to commit themselves to making future payments. The main questions to be addressed are whether in equilibrium $b' > 0$, and if so, how large it is.

For the same reasons as in the previous section, W should be set equal to $2w_1$ unless that would violate the constraint on W; in which case the second-period wage is just enough to prevent mobility of workers who do not need to move. Explicitly,

$$(4) \qquad W(m_2) = 2w_1, \qquad m_2 \leq 2w_1 - b',$$
$$= m_2 + b', m_2 \geq 2w_1 - b'.$$

Granted (4), the derivative of expected utility with respect to b' is

$$(1-p)\int_{2w_1-b'}^{\infty} U'(m+b')dF(m) + pEU'(m'+b'),$$

where F is the distribution function for m_2. The derivative of the first constraint in (3) (i.e., minus profit per man) with respect to b' is

$$(1-p)[1-F(2w_1-b')] + p = 1-(1-p)F(2w_1-b').$$

Since the value of the constraint is $u'(w_1)$ (by variation of w_1), the first-order condition for b' is that

$$(5) \qquad (1-p)\int_{2w_1-b'}^{\infty} U'(m+b')dF(m) + pEU'(m'+b')$$
$$- [1-(1-p)F(2w_1-\mathrm{b}')]u'(w_1) = 0,$$

with $b' > 0$; or that, alternatively, $b' = 0$ and the left-hand side of (5) is nonpositive. When $b' > 0$, first-period marginal utility equals second-period expected marginal utility.

Using (4), we also have the zero profit condition:

$$(6) \qquad w_1 - m_1 + (1-p)\int_0^{2w_1-b'} (2w_1 - b' - m)dF(m) + b' = 0.$$

That is, first-period wages are below the first-period marginal product to finance a second-period unconditional transfer of b' and an increase in the second-period wage when (observed) marginal product is low. When can (5) and (6) be satisfied simultaneously with $b' > 0$? One case is where the endogenous mobility threat would not be effective with $b' = 0$ ($m_2 < 2w_1$ for all m_2).[3]

Proposition 1: Suppose that m_2 is known never to take a value greater than

$$(7) \qquad M = \frac{m_1 + (1-p)E(m_2)}{\frac{3}{2} - p}$$

and that

$$(8) \qquad U'(M) < EU'(m').$$

Then the optimal b' is positive.

Proof: We show that, under the stated assumptions, when b' is set to zero in (6), the left-hand side of (5) is positive. It is therefore impossible that $b' = 0$, and the desired conclusion follows. With $b' = 0$, it is readily seen that the solution of (6) is

$$w_1 = M/2,$$

for in that case $F(2w_1) = 1$ and the integral is $M - E(m_2)$. That is, with zero transferable pension it is feasible to equalize pay in both periods.

Turning to the left-hand side of (5) with this value of w_1, we see that the integral vanishes, and the expression in square brackets becomes p. We get

$$pEU'(m') - pu'(M/2) = p[EU'(m') - u'(M)],$$

which is positive by assumption (8). Q.E.D.

It will be noticed that, when U' is convex, a sufficient condition for (8) is that $M > E(m')$. Granted the first assumption, this is eminently plausible, as indeed is (8) for any utility function. The significance of the result is that $b' = 0$ only if the distribution of m_2 is highly dispersed or departure under "necessity" tends to lower the marginal utility of consumption. The upper bound M is quite restrictive: with $m_1 = E(m_2)$ and $p = 1/2$, $M = (3/2)E(m_2)$; but the conditions of the proposition are quite strongly sufficient. We may safely conclude that often $b' > 0$, recognizing that necessity is likely to raise the marginal utility of consumption.

It is not so easy to see just how large b' is likely to be. The greater it is, the more commonly will second-period wages and the final pension be greater than first-period wages. Some insurance for exogenous mobility is available, but at the expense of reduced insurance against low second-period productivity when mobility is unnecessary.

The case where m' is always zero is that of disability, and in that case b' is positive whether or not m_2 is bounded above by M. We have studied this case (with m_2 nonrandom) in a previous paper from the point of view of the social optimum. We shall return to the many-period disability case below in section 11.5 where we address the commitment issue in more detail.

11.4 Endogenous Mobility

We now study a version of the model in which workers may have different values to different employers, the values not being known in the first period. Because workers cannot bind themselves in advance to remain with their first employer in the second period, contracts must be negotiated on the assumption that a worker will move if it pays him to do so.[4] It will pay him to do so if a wage offer from some other employer (net of costs of moving) plus any transferable pension rights exceed the wage (plus pension) available if he stays with his first employer. To emphasize the contrast with the case of exogenous mobility, we suppose that there is no utility or disutility attached to moving, other than what can be allowed for in the alternative wage offer; and indeed we assume, in the first instance, that this net wage offer is observable. This observability assumption is unduly strong for some cases. We also consider two other alternatives: that the wage offer is not at all observable and that the wage offer is observable for the transferable pension (which is not paid until period 3), but not observable at the time of wage offer from the first-period employer. As in the implicit contract literature (see Hart [1983] for a survey) productivity with the first employer in the second period may not be observable. We deal with that by considering two cases, depending on whether m_2 is or is not observable. Thus we have six cases to analyze, of which we look at five. We do not consider more realistic, less extreme cases where outside offers are sometimes observable and marginal product is imperfectly observable.

Marginal productivities with the original firm and elsewhere (identified with the observable wage offer) m_2 and m_2' are jointly distributed without atoms, the distribution being given by a density function,[5] $f(m_2, m_2')$. Thus, we assume that the distribution of m_2' is independent of the wage contract. With both variables observable, W and b' are functions of both m_2 and m_2'. The commitment constraint is that b' be nonnegative, and movement is determined by which of W and $m_2' + b'$ is the greater. Equilibrium is given by

$$(9) \qquad \max_{w_1, W, b'} \quad u(w_1) + EU[\max(W, m_2' + b')]$$

subject to

$$w_1 - m_1 + \iint_{W \geq m_2' + b'} (W - m_2) f dm_2 dm_2' + \iint_{W < m_2' + b'} b' f dm_2 dm_2' = 0,$$

$$b' \geq 0.$$

As before, the shadow price of the constraint is $u'(w_1)$. For each m_2 and m_2', W and b' have to be chosen optimally, that is, so that the function

$$(10) \qquad V(W,b') = \begin{cases} U(W) - u'(w_1)(W - m_2), & W \geq m_2' + b', \\ U(m_2' + b') - u'(w_1)b', & W < m_2' + b', \end{cases}$$

is maximized subject to $b' \geq 0$. When $m_2 > m_2'$, the first of these expressions is larger than the second when we set $W = m_2' + b'$. Therefore we maximize the first expression subject to the constraint $W \geq m_2'$ (which follows from $b' \geq 0$). This is achieved by setting W equal to the larger of $2w_1$ and m_2'. In that case, b' can be anything between zero and max $(2w_1 - m_2', 0)$, since it will not be collected. When $m_2 < m_2'$, a similar argument shows that the worker will be induced to move. Then b' is set equal to $\max(2w_1 - m_2', 0)$, and W takes any value less than $\max(2w_1, m_2')$ since it will not be collected.

Summarizing, we have

Proposition 2: With endogenous mobility, observability of wage offers and marginal products, and no precommitment by workers, workers go where their productivity is greatest in the second period. If productivity is greatest with the original employer, income in the second and third periods is w_1 or $m_2'/2$, whichever is the greater. If the best alternative wage offer is greater than productivity with the original employer, the transferable pension is just sufficient to increase total income in the second and third periods to w_1, if that is greater than $m_2'/2$, and the transferable pension is zero otherwise. The proposition is illustrated in figure 11.1.

With m_2 not observable by the worker, the firm guarantees minimal levels of W and b' and preserves the option of raising either variable when m_2 is observed by the firm, if that is in the interest of the firm. We denote the guaranteed levels by $W_0(m_2')$ and $b_0'(m_2')$. The worker is assumed to know the distribution of payments the firm will actually make. Thus the equilibrium contract remains the solution to (9) with the additional control variables W_0, b_0' and the additional constraints

$$(11) \qquad \underset{W,b'}{\max}\; V^*(W,b') = \begin{cases} m_2 - W, & W \geq m_2' + b', \\ -b', & W \leq m_2' + b', \end{cases}$$

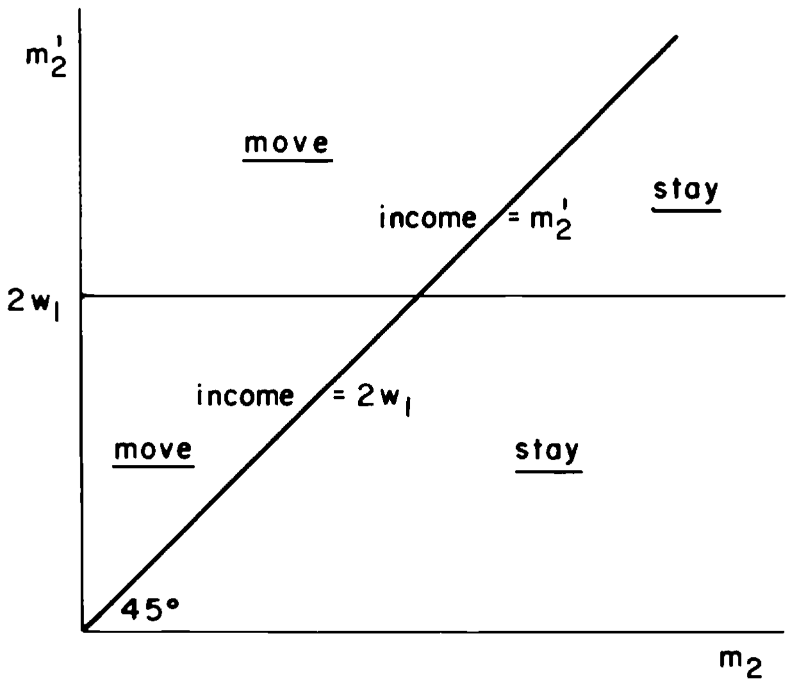

Fig. 11.1 Propositions 2 and 3.

subject to $$W(m_2,m_2') \geq W_0(m_2')$$
$$b'(m_2,m_2') \geq b_0'(m_2').$$

The firm can precisely duplicate the optimal contract when m_2 is observable by setting W_0 equal to $2w_1$ and $b_0'(m_2')$ equal to $\max[(0, W_0 - m_2')]$ and choosing W and b' to induce mobility if and only if $m_2' > m_2$. Since the contract in this case satisfies a maximization problem without the additional constraints, the optimum with observability is the optimum without observability. Thus we have

Proposition 3: With outside wage offers observable, the optimal allocation is the same whether marginal productivity at the original firm is observable or not.

The outcome is rather like the Harris-Holmstrom case, with the "wage," or income, remaining constant unless the alternative employment opportunity makes an increase in income necessary. It is not the unpredictable increase in the worker's marginal productivity to the economy that brings about an (undesirable) increase in his income, but the unpredictable increase in his marginal productivity to the rest of the economy. Thus an economy with fewer but larger firms should have an equilibrium closer to the ideal case, since there are fewer possibilities outside the firm, implying that the wage then increases in fewer states of nature and a higher initial wage results in equilibrium.

If workers leave one employment in search of a higher income, or have incentive to show a low wage ($w_1 > m_2/2$, $m_2' > m_2$), or are—less realistically—unable to prove that alternative wage offers are genuine, W cannot be made to depend on m_2', although, perhaps with the help of the State, b' might still depend on m_2'. We examine next the case where b' is a function of m_2' but W is not. We also make the simplifying assumption (ignoring the search case) that the worker knows m_2' when he decides whether to move or not, although m_2' is not verifiable by the employer.

If m_2 is observable, W depends only on m_2 but b' can depend on both m_2 and m_2', then, considering a particular m_2, W, and b', maximize

$$\int V(W, b') f \, dm_2', \tag{12}$$

with V defined (as in [10]) by

$$V(W, b') = \begin{cases} U(W) - u'(w_1)(W - m_2), & W \geq m_2' + b', \\ U(m_2' + b') - u'(w_1)b', & W < m_2' + b'. \end{cases}$$

We should note straight away that, in this case, W is never less than $2w_1$, for if it were we could increase V by increasing W and simultaneously increasing b' if necessary to keep the division between the two cases in the definition of V unchanged.

Now consider the choice of b', keeping W fixed for the moment. b' is constrained to be nonnegative. Therefore $m_2' > W$ implies $W < m_2' + b'$, and, bearing in mind that $W \geq 2w_1$, b' should be made as small as possible:

(13) $$b' = 0, m_2' > W.$$

Suppose now that $m_2' \leq W$. If $m_2 \geq m_2'$, the worker should be induced to stay, for $U(W) - u'(w_1)W \geq U(m_2' + b') - u'(w_1)(m_2' + b')$ when $W > m_2' + b'$. Thus

(14) $$b' < W - m_2', m_2' \leq m_2, W.$$

If $m_2 < m_2'$, the worker should be induced to move, with wage plus pension as close to W as possible. This can be done by setting

(15) $$b' = W - m_2' + \epsilon, m_2 < m_2' \leq W,$$

where ϵ is a small positive number. ϵ should be "infinitesimally" small.

With these choices of b', we have

$$V(W, b') = \begin{cases} U(W) - u'(w_1)(W - m_2), & m_2 \geq m_2' \leq W, \\ U(W) - u'(w_1)(W - m_2'), & m_2 < m_2' \leq W, \\ U(m_2') & m_2' > W. \end{cases}$$

Here, since we have taken account of the mobility decision induced by b', we can set $\epsilon = 0$. When $m_2 \leq 2w_1$, W should be set equal to $2w_1$, since it maximizes

$$\int_0^W [U(W) - u'(w_1)W] f\, dm_2' + \int_W^\infty [U(m_2') - u'(w_1)m_2'] f\, dm_2'.$$

When $m_2 > 2w_1$, W should be less than m_2; for when $W > m_2$, the integral of V is

$$\int_0^W [U(W) - u'(w_1)(W - m_2')] f\, dm_2' + \int_W^\infty U'(m_2') f\, dm_2'$$
$$+ u'(w_1) \int_0^{m_2} (m_2 - m_2') f\, dm_2',$$

which is increased by reducing $W > 2w_1$. Therefore in fact W maximizes

$$\int_0^W [U(W) - u'(w_1)(W - m_2)] f\, dm_2' + \int_W^\infty u(m_2') f\, dm_2'.$$

The first-order condition is

(16) $$[U'(W) - u'(w_1)] \int_0^W f\, dm_2' = u'(w_1)(W - m_2) f(m_2, W)$$

We have proved

Proposition 4: With endogenous mobility, observability of marginal products with the original firm, but observability of wage offers only for the determination of transferable pensions, a worker whose second-period productivity in all firms is less than twice his first-period wage has the same income in all periods of life, and works where his productivity is greatest; while one whose second-period productivity is greater than that in some employment has no transferable pension if he moves, and if he does not move has a wage plus pension less than his productivity, given by (16).

This situation is illustrated in the figure 11.2.

Equation (16) does not tell us much about W without further assumptions. When m_2 and m_2' are independently distributed, $f(m_2, m_2') = g(m_2)g'(m_2')$, and we have

$$m_2 = W + \left[1 - \frac{U'(W)}{u'(w_1)}\right] \frac{\int_0^W g'\,dm_2'}{g'(W)} \tag{17}$$

If, as is reasonable, $\int_0^W g'\,dm_2'/g'(W)$ is an increasing function of W for W

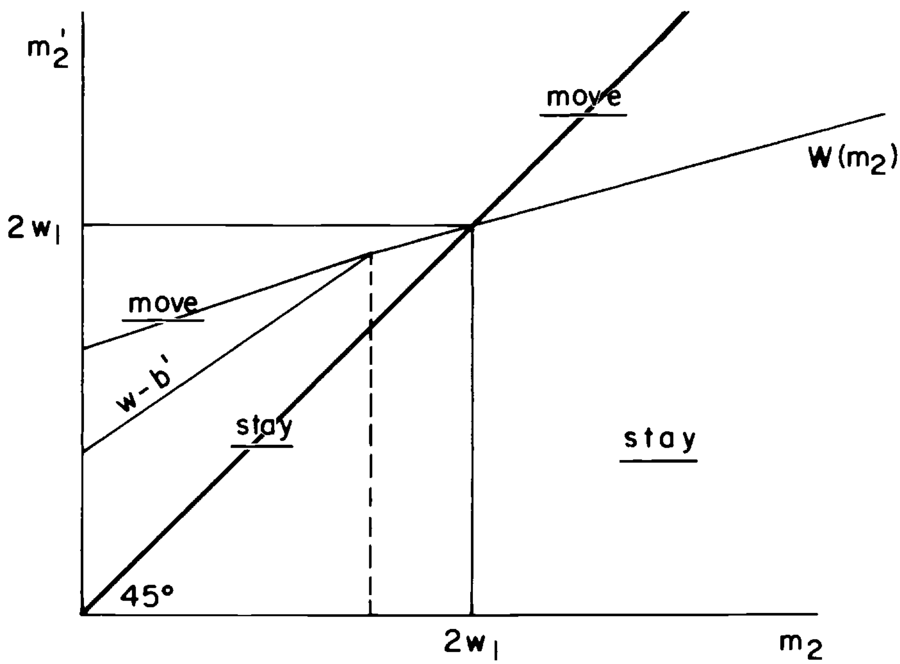

Fig. 11.2 Proposition 4.

$\geq 2w_1$, the right-hand side is an increasing function, with gradient greater than one. It follows that W is an increasing function of m with gradient less than one.

If employment contracts do not condition b' on m_2, but do condition it on m_2', the results are a little less neat, and we do not explore them here.

With m_2 observable, equilibrium is given by the same maximization, (9), except that W and b' are now constrained to be independent of m_2'. Before turning to the first-order conditions, let us note that W and b' are set to induce some mobility provided there is a positive probability that m_2' exceeds m_2. There is no mobility if $W - b'$ exceeds the maximal possible m_2'. In this case, consider raising b' to any level not exceeding $W - m_2'$. Doing this makes those who move better off and saves revenue for the firm, since each mover represents a loss in productivity that is no greater than the savings in compensation from mobility. Of course, this argument does not imply that the optimal b' equals $W - m_2$. Thus we have $W - b' \leq \max m_2'$. The usual envelope argument implies that this inequality is strict for the optimal b'. This being so, the first-order conditions are

(18)
$$[U'(W) - u'(w_1)] \int_0^{W-b'} f dm_2' - u'(w_1)(W - b' - m_2)f(m_2, W - b') = 0$$

(19)
$$\int_{W-b'}^{\infty} [U'(m_2' + b') - u'(w_1)] f dm_2' + u'(w_1)(W - b' - m_2)f(m_2, W - b') = 0$$

or (19) is nonpositive with $b' = 0$.

We analyze the two cases separately. Assume $b' > 0$ for the moment. The integral in (19) is taken over the range $m_2' > W - b'$. Thus we have

$$U'(m_2' + b') < U'(W).$$

Using this in (19), we find that

$$[U'(W) - u'(w_1)] \int_{W-b'}^{\infty} f dm_2' + u'(w_1)(W - b' - m_2)f(m_2, W - b') > 0.$$

Comparing this with (18), it is found that

(20)
$$U'(W) - u'(w_1) > 0$$
$$W - b' - m_2 > 0.$$

From (20) we have

(21)
$$2w_1 > W > m_2 + b' > m_2.$$

Thus b' is necessarily zero for $m_2 \geq 2w_1$. The latter condition in (20) represents an implicit tax on moving (workers who stay have net compensation more than their marginal products).

Now assume $b' = 0$. Then, (18) becomes

$$[U'(W) - u'(w_1)] \int_0^W fdm_2' = u'(w_1)(W - m_2)f(m_2, W).$$

Thus we have

$$\text{sign}\,(W - 2w_1) = \text{sign}\,(m_2 - W).$$

Thus W lies between $2w_1$ *and* m_2 for values of m_2 for which b' is set equal to zero. We can conclude that workers with high marginal products are paid less than their marginal products.

Summarizing these results, we have

Proposition 5: With endogenous mobility, observability of marginal products with the original firm but no observability of wage offers, and no precommitment by workers, income is lower in the second and third periods than in the first for any worker who remains with the original employer and has marginal product below the first-period wage. For workers with positive transferable pensions, the pension is les than $W - m_2$. For any worker who remains with the original employer and has marginal product above the first-period wage, there is no transferable pension, and income is higher in the later period than in the first period but less than the marginal product. The question of which values of m_2 have b' positive appears to be difficult in general, depending on the properties of f. The proposition is illustrated in figure 11.3.

One case remains to be analyzed, that where neither m_2 nor m_2' is observable. In this case, the firm chooses w_1, W_0, b_0', $W(m_2)$, and $b'(m_2)$ to maximize (9) subject to the constraints of self-interest in the second period:

$$\max_{W,b'} \int_0^{W-b'} (m_2 - W)fdm_2' + \int_{W-b'}^{\infty} -b'fdm_2' \tag{22}$$

subject to $W \geq W_0$

$b' \geq b_0'$

Thus we have the first-order conditions

$$\begin{aligned} &(m_2 - W + b')f(m_2, W - b') - \int_0^{W-b'} fdm_2' \leq 0 \\ &-(m_2 - W + b')f(m_2, W - b') - \int_{W-b'}^{\infty} fdm_2' \leq 0 \end{aligned} \tag{23}$$

with equality if $W > W_0$ and $b' > b_0'$ respectively. From the signs of the different terms, we have either $W = W_0$ or $b' = b_0'$, or possibly both.

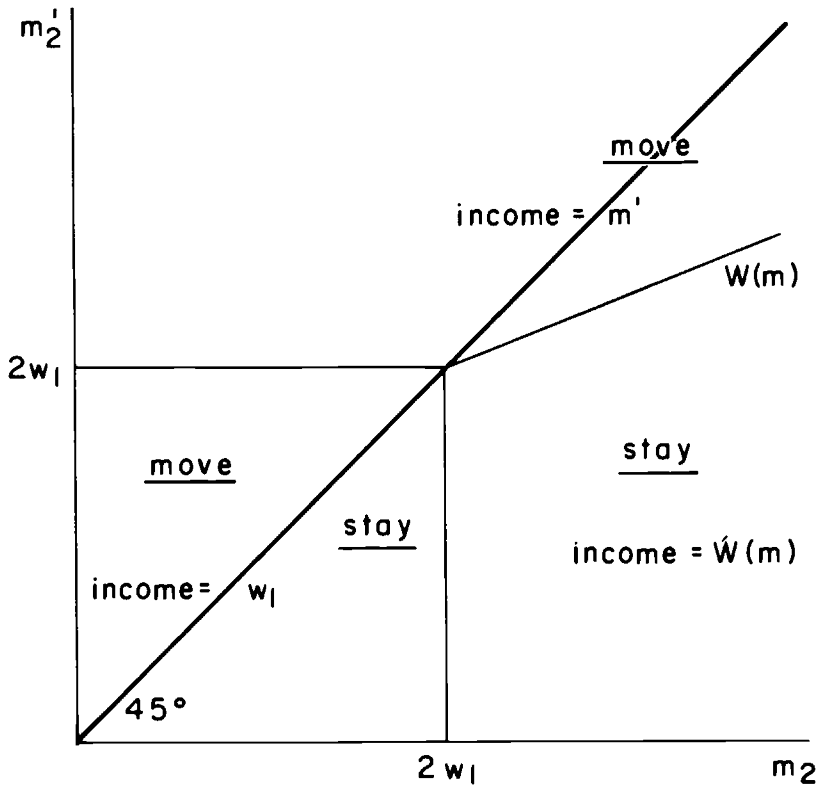

Fig. 11.3 Proposition 5.

Moreover, $b' = b_0'$ for $m_2 > W_0 - b_0'$ and $W = W_0$ for $m_2 < W_0 - b_0'$. In addition, where $W > W_0$, $W < m_2 + b_0'$. Where $b' > b_0'$, $b' < W_0 - m_2$.

To analyze W_0 and b_0' we would need to consider the first-order conditions from (9). This is a messy calculation since W depends on b_0' when $W > W_0$ and b' depends on W_0 when $b' > b_0'$, and we leave analysis of this problem for the future.

11.5 Early Retirement

In the models analyzed above, everyone worked in the first two periods and no one worked in the third. This approach ignores the wide spread in the distribution of retirement ages that is currently observed. Indeed, many pension plans allow a choice of retirement age and relate the size of pension to the age at which it is first claimed as well as wage history and job tenure.

Our earlier analyses of public retirement systems concentrated on the relationship between optimal benefit size and age at retirement. In that work, there is naturally no constraint corresponding to the nonnegativity of transferable pension rights. We recall the ideas of that analysis briefly and relate it to the issues in the current paper by checking when the nonnegatively constraint on worker debt to the firm needed for private pensions is binding.

Assume that both the productivity of workers and the fact of working are observable but the ability to work (taken to be a zero-one variable) is not. Then, ignoring those unable to work their entire lives, one can implement a system of lump sum redistributions based on productivity and have actuarially fair pensions. Actuarially fair pensions imply that the ex post lifetime budget constraint varies with the length of working career in the same way as output produced. Length of working career depends on both choice and random factors affecting productivity. (This is a shorthand for factors that affect the disutility of work and the availability of jobs as well as productivity.) There are two reasons for believing that one may be able to improve on the choice of actuarially fair pensions for a public redistribution and retirement system—income redistribution and insurance provision.

To isolate the income redistribution motive, let us temporarily ignore the uncertainty about work ability. If individuals differ in (unobservable) labor disutility as well as productivity, then lump sum redistribution based solely on productivity still leaves room for further redistribution based on disutility of labor. Normally, those retiring earlier will have a higher social marginal utility of consumption. It is therefore worthwhile to distort the retirement decision to redistribute further toward early retirees, who have lower ex post budget constraints. (Essentially the same argument holds when we consider the insurance argument below.) This argument is strengthened once we recognize that redistribution for productivity differences is done by a distorting annual income tax rather than a lump sum tax. Income redistribution is then limited by work incentive problems on hours worked. Rather than an actuarially fair pension system (with fairness defined in terms of the government's total budget), it is likely to be appropriate also to redistribute in a way that discourages longer working lives, once we recognize the positive correlation between productivity and length of working life (which reflects a negative correlation between productivity and disutility of labor as well as higher compensation).

Under conventional economic assumptions redistribution plays no role in private pension design (although unions may be concerned). Indeed, adverse selection problems imply that the attempt to provide insurance against early loss of earnings ability will have a cost. We ignore both redistribution and adverse selection in the rest of this section.

With all individuals ex ante identical in productivity, preferences, and disability risk, actuarially favorable early retirement provisions provide insurance against an early loss of earnings ability and are a natural part of the optimal labor contract. In our earlier papers, we derived equations for the optimal wage and retirement benefit plans under the alternative assumptions that workers do, or do not, save on their own. If we now make the same commitment assumptions as in the first part of the present paper and assume identical opportunities elsewhere, the socially optimal contracts remain privately feasible as long as expected future compensation is not less than expected future productivity. We now examine when this condition is satisfied.

In both of our earlier papers we related lifetime consumption under the optimal plan to length of working life for the case of constant marginal product. Lifetime consumption was an increasing function of working life, but increasing more slowly than lifetime production (until the planned retirement date, when the rates of increase are equal). Put differently, the optimal plan has an implicit tax on work throughout the working life up to the planned retirement age, when the tax is zero. This is the natural way to provide insurance against the adverse event of early retirement. Under the simple institution of a wage paid that is either consumed or saved, the optimal plan is thus not sustainable under mobility since the taxes on work are used to finance a lump sum transfer at the start of working life.

Pensions, however, represent a different institutional setting from simple annual wage payments, with pension benefits at a future date depending on the date of retirement. In actual pensions, benefits are paid conditional on stopping work at one's own firm. In a world of uniform marginal products and costless mobility this condition has no bite, and pensions do not offer any greater insurance possibilities than simple wages. If pensions are conditional on full retirement, then there is the possibility of insurance. The private equilibrium can support the social optimum if there is implicit taxation of early work, but not if there is implicit subsidization of early work. To examine the feasibility of privately imitating the social optimum we consider the following institution. Work at age t results in a wage of $c(t)$ which is fully consumed. Retirement at age t results in a benefit $b(t)$ for all later periods of life. Mobility at age t results in a lump sum transfer of resources $R(t)$ which is taken to an alternative firm to help finance retirement. The socially optimal system is privately supportable if the optimal $b(t)$ and $c(t)$ imply a net surplus of output over expected payments, which is nonnegative.

A worker who moves at age t must bring to his new employer an amount sufficient to cover the expected value of future payments to him, net of his product in the firm, in case he stays with a new employer until retirement. If this is so when he stays, it is also the case when he moves again, provided that the new transfer sum is similarly just sufficient to cover the expected net loss from employing him. We express this formally, using the notation of our previous papers, where $F(t)$ is the probability that a worker is still able to work at age t; r is the age at which a healthy person retires; T is the length of life; and m is the constant marginal product. It is convenient to write the equation in terms of the expected values of transfer, and of payments subsequent to t, with the expectation taken as of time O. The same result of course follows if we use conditional probabilities. The required level of $R(t)$ satisfies

$$(24) \qquad R(t)[1 - F(t)] = \int_t^r [c(s) - m][1 - F(s)]ds + \int_t^r b(s)(T - s)f(s)ds + b(r)(T - r)[1 - F(r)].$$

R would be zero at $t = 0$ by the budget balance constraint for the social optimum. R is evidently positive at $t = r$. If the right-hand side should be first a nondecreasing function of t, then nonincreasing, it would follow that $R \geq 0$ for all t. Differentiating, we see that this is true when there exists $s \leq r$ such that

$$(25) \qquad c(t) + b(t)(T - t)\frac{f(t)}{1 - F(t)} \quad \begin{matrix} \leq m \text{ for } t \leq s \\ \geq m \text{ for } s \leq t \leq r. \end{matrix}$$

Since, as we showed in the papers referred to, b and c are increasing functions of t, a sufficient condition for $R \geq 0$ is that

$$(26) \qquad \frac{(T - t)f}{1 - F}$$

is a nondecreasing function of t.

This holds in particular for the uniform distribution, and more generally when

$$(27) \qquad f(t) = \frac{ae^{at}}{e^{aT} - 1} \quad a > 0.$$

Since (26) is very strongly sufficient for (25), we conclude that it is none too easy to find simple cases where the social optimum should not, in theory, be privately implementable. Yet again, we find that a system of transferable pension rights is, besides being evidently desirable, consistent with private rationality. Even if the socially optimal R were sometimes negative, it would still be true that a constrained optimum would have R positive for ages of mobility.

11.6 Inflation

The analysis above made no mention of inflation. If inflation were fully neutral, the analysis above would stand and imply labor contracts in real terms. However, inflation often occurs at times when other factors are changing too. To analyze inflation in the setting used above, one would want to know how the joint distribution of marginal products with and without mobility tends to change as inflation rises. Also one would want to incorporate changes in real interest rates. We have not pursued such an analysis.

It is natural, also, to ask how existing pension structures are affected by inflation. This requires consideration of the full labor contract. With a defined contribution pension plan, workers bear the real interest rate risk associated with inflation assuming that the response of wages to inflation is independent of the size of existing pension fund. While this is a plausible assumption, there is nothing to prevent actual or implicit contracts from relating current wages to pension fund performance. Analysis of optimal sharing of real interest risk would require a description of the full portfolio positions of workers and shareholders. Under a defined contribution plan, workers who have left a firm bear the full risk.

With defined benefit plans, we should again distinguish between departed and present workers. For departed workers, the presence of a promise to pay future sums stated in nominal terms is a plan to make workers worse off, and firms better off, the greater the inflation rate. Decreases in the real rate of return would lessen the advantage accruing to firms. Only when a rise in inflation decreases the nominal rate of return does it move the utility of firms and workers in the same direction. Thus defined benefit plans appear, at first examination, not to be part of an optimal contract.

For workers staying with a firm, the sharing of inflation risk depends on the relationship of final wages to inflation. In some cases it is clear that workers are protected as part of the long-term contract (in contrast to the lack of protection under spot contracts, as analyzed by Bulow [1982]). For example, if wages depend on job done and on age, and the allocation of workers to jobs depends on ability, and perhaps seniority (with the usual effect of seniority), pension obligations are in real terms to the extent that labor contracts will be independent of past pension obligations. This independence will generally depend on the proportions of workers of different ages and is very likely to be true where near retirees are a small fraction of the labor force.

11.7 Pension Plans

The analysis so far has considered a single worker. In effect, it applies to pension arrangements for a body of workers who are ex ante identical.

Even if workers differ in some dimension but behave identically in supplying labor to the firm, the analysis applies. We can consider the discovery by workers of their true productivity distributions as one of the risks being insured by the labor contract. The only problem for the firm is having the correct productivity distribution for doing the expected present discounted value calculation.

If workers are aware of their differences (or merely have labor supply responses to contractual terms that are correlated with productivities), then there is an adverse selection problem along with the moral hazard problem of mobility. (We continue to ignore other moral hazard problems associated with effort.) We do not model this formally, but simply ask what kind of worker is particularly attracted to the equilibrium contract for a given level of expected productivity. We suppose that alternative employments offer wages equal to marginal products each period.

When there is no equilibrium mobility, and wages are downwardly rigid, the convexity of the second-period wage schedule in terms of productivity (eq. [2]) makes the contract particularly attractive to high-risk workers, as well as to risk-averse workers. In the case of endogenous mobility, workers with lower anticipated outside offers are particularly attracted. These workers are most likely to collect on the insurance premium implicitly paid in the first period. In the case of exogenous mobility, the answer depends on the division of insurance benefits between unfortunate moves and low productivity in continued employment. In the mixed case, the latter seems likely to predominate empirically.

Once we consider heterogeneity of the labor force, two of our simplifying assumptions—no worker savings and risk neutrality of the firm—appear much less satisfactory. Without savings by workers, there is considerable symmetry between oversaving and undersaving for a particular worker's ex post (and so ex ante) position, as can be seen by considering movements along an intertemporal budget line away from the optimal allocation. With worker savings we introduce a further moral hazard problem (as has been discussed in our earlier papers). In addition, any asymmetries in market opportunities for savings and dissavings would imply that workers find it easy to save to offset undersaving but hard to borrow to offset oversaving. This asymmetry translates into an asymmetry in the evaluation of pension plans that oversave for some and undersave for others. Presumably, that is a case for smaller plans than the analysis above would suggest, but we have not undertaken formal analysis.

Risk neutrality of firms was a very handy simplifying assumption. But bankruptcy means that firms are not risk neutral either about outcomes or about paying promised benefits in all states of nature. We do not pursue this. We have also ignored the complications that come from firm-specific risks and decreasing returns, which have been central to much of the implicit contract literature.

11.8 Repeated Negotiations

The models analyzed above can be used for normative purposes; for example, one can evaluate the degree of success of existing arrangements in achieving a constrained optimal allocation for the models. With the additional assumption of optimal contracting, the model becomes a positive model of equilibrium. As a positive model, some of its assumptions may be too inaccurate empirically to serve as an adequate model. One such assumption is that of objective probabilities. There is little reason to believe that subjective probabilities about events affecting the value of the lifetime labor contract are either similar between worker and firm or correct for workers. Firms have a direct financial experience to draw on in revising beliefs. Workers experience retirement a small number of times. Insofar as subjective beliefs of workers and firms are different, labor contracts will, in part, be bets on the events for which the parties have different beliefs. Thus, inaccuracy of the assumption of identical beliefs is likely to have sizable implications for the equilibrium allocation of resources.

A second source of potential shortcoming is the assumption of a lifetime contract that is binding on the firm. Mostly, we see labor contracts that are at-will contracts (no set termination, with freedom to terminate at any time or after notice which is short relative to a working lifetime) or contracts for a relatively short time, one to three years. The short contracts, however, are associated with considerably longer relationships, on average.[6] Since circumstances generally change over time, there is little reason to think that a series of equilibrium short-term contracts (under some bargaining theory) would produce the same allocation as a single long-term contract (under the same bargaining theory). However, this may not be the right comparison if a model of short contract equilibrium is not an accurate picture of the economic environment. We mention three reasons for doubting that account. First, some agents behave as if they were legally bound even when they are not. Second, firm reputation may affect both current worker efficiency and the ability to attract additional workers, making it optimal for a firm to carry out implicit contracts. We return to a reputation model below. Third, government rules affecting the labor market (both legislated and common law) affect the validity of the assumptions of the classical short-term contract equilibrium model (i.e., imply some ability to commit beyond the length of contract). We will discuss these mechanisms in the process of discussing the short-term model introduced by Bulow and coauthors (Bulow 1982; Bulow and Scholes 1984; Bulow et al. 1984).

In the classical competitive model the wage equals the marginal product of labor. If part of the wage is paid currently and part in deferred compensation, total compensation in equilibrium equals the marginal product. Presumably, the fact that equilibrium is defined in terms of a relationship between total compensation and marginal product would generalize to a matching model where the next best alternatives for both firm and worker were strictly inferior to their current match.[7] Then total compensation would continue to be determined independently of the pension structure. Under this interpretation (and a finite horizon), one cannot have a perfect equilibrium where firms carry out implicit contracts that imply total compensation in excess of marginal product late in life. Presumably, in reaching the short-term contract equilibrium, some further mechanism (such as transaction costs) is involved to rule out explicit longer-term contracts that provide insurance.

For both of these models (classical and matching), we can examine the robustness with respect to legal and institutional considerations of the result that the absence of explicit long-term contracts implies a short-term contract equilibrium in terms of total compensation. The Age Discrimination Act prohibits discrimination on the basis of age. Generally, nondiscrimination legislation has been interpreted in terms of wages, not total compensation, making it illegal to pay lower wages to an otherwise identical worker for whom fringe benefits are more expensive. Presumably this would hold for defined benefit pensions, where the cost of the pension varies with both age and seniority. Thus in the presence of this act, we cannot have an equilibrium with both defined benefit plans and age uniform total compensation.

Even without the Age Discrimination Act, common law requires good faith exercise of discretion in completion of a contract (see Burton 1980). Good faith can be seen as a preservation of expectations. Thus, if everyone understands that equilibrium is in terms of total compensation, it is not bad faith to carry it out. If workers believe that wages are independent of pension promises in "normal" times and pension plans will not be terminated in "normal" times, then firing of workers, plan terminations, and low wage offers solely to preserve a relation between total compensation and marginal product would probably be held to be bad faith actions and result in liability for compensatory damages.[8] Thus, legal restrictions can enforce and make credible some implicit contracts much as they do explicit contracts.

Thus, it appears that legal interpretations of employment relations and pension provisions would sustain an equilibrium with some degree of insurance for workers even in an economy where explicit intertemporal labor contracts did not exist. In addition to this mechanism, it may be possible to construct a perfect equilibrium with implicit contracts once one recognizes that firms have (potentially) infinite horizons.

Although it seems to be a promising line of inquiry, we have not developed a formal model of equilibrium (perfect or otherwise) with reputation effects. Nevertheless, we shall sketch briefly how such a model might look and what it might imply.

To start, assume no uncertainty for a firm but uncertainty for individuals plus a desire to save by way of pensions. Assume that the supply of new workers depends on current treatment of older workers as a basis for forecasting how new workers will be treated in the future. (This overlapping generations model seems simpler than the probably more important effect of reputation on efficiency of existing workers.) In some circumstances, it becomes worthwhile for firms to overpay (at least some) current older workers relative to their marginal products and alternative jobs. In a steady state, with naive expectations and an interest rate equal to the growth rate, the equilibrium contract will probably be similar to the ones analyzed in sections 11.2–11.4 above. The critical condition is that the profitability of continuing to fulfill these contracts exceeds that from exploitation of current older workers and an end to implicit contracting.[9] Critical to this possibility is the assumption of noninsurable, nondiversifiable risk, which makes a risk-averse worker value the asset of future benefits more highly than the firm values the liability of providing those benefits. In this equilibrium, one will have inflation protection for workers who have pension benefits keyed to final wages or pension benefits regularly adjusted in step with general wages. One may also have an effect on capital accumulation similar to that of unfunded social insurance.

The use of reputation rather than contracts for this mechanism affects the outcomes in a major way once one recognizes the importance of uncertainty for firms. First, events may occur that decrease the value of maintaining reputation, for example, bankruptcy or a corporate takeover permitting a repudiation of the previous management's implicit contracts. Then, the optimal contract must consider not only the problem of worker mobility but also the need to maintain the value of preserving reputation. This fact probably limits the extent to which insurance can be offered to workers and brings in a major role for funding rules in determining sustainable benefit packages. Second, a model with firm risk probably implies the sharing of profitability risk by workers, a phenomenon that appears to be present in the United States.[10] Pension rules then affect the distribution of wage risks among workers of different ages. There is an a priori suspicion that final wage defined benefit plans put too much risk on near retirees because of the leverage of wage changes on future pension benefits. The same conclusion would hold for a flat or service-related plan that sustained a constant ratio of pension benefit parameter to wages. Third, the implicit contract sustained by reputation effects is likely to be a very complicated contract, suggesting all the concerns normally associated with consumer protection in the presence of complicated contracts.[11]

11.9 Summary

Using a variety of competitive three-period models, this paper analyzes equilibrium labor contracts. It is assumed that risk-neutral firms save for workers and try to provide insurance against poor levels of second-period productivity. The first period has known productivity, while the third has retirement. With no restrictions on observability or commitment (and zero interest, for convenience), the optimal contract would yield constant consumption over the three periods, equal to one-third of the expected value of the sum of productivities in the two working periods. In the second period, the worker would move to another firm if and only if the offered wage exceeded his marginal product with his first-period employer. The worker would receive a payment from the first-period employer if his second-period wage were less than the contracted consumption level over his two remaining periods, and the worker would make a payment to his first-period employer if his second-period wage exceeded the contracted consumption level over his two remaining periods. The starting place of our analysis is that this contract is not attainable. Throughout the paper, we assume that the worker cannot commit himself either to staying with his first-period employer in the second period or to paying compensation to his first-period employer should he leave the firm. We assume that the firm can commit itself to a three-period contract, limited only by observability restrictions. (We also speculate on how the analysis would change if firm reputation replaced commitment.)

If it is known that the worker's productivity with his first-period employer is the same as his best alternative wage offer, there is a threat of mobility limiting the contract, but no reason for the worker to leave the firm that helps the design of the optimal contract. The optimal contract has nondecreasing consumption over time and increasing consumption if second-period productivity is large enough. In order to limit the extent to which the mobility threat curtails the provision of insurance, the optimal contract has no further payment to any worker who leaves his first-period employer. When there is no social reason for mobility, it is natural for the contract to discourage mobility as much as possible—by having no transferable pension, in this case.

The central focus of the paper is the examination of circumstances where this conclusion does not hold. That is, we identify circumstances where the optimal contract contains deferred payments to workers who switch employers.

First we note that workers sometimes leave firms for reasons other than higher wages elsewhere. These reasons include health and interpersonal relations which limit the ability to continue performing the work (or raise its disutility) and a variety of reasons for geographic mobility that may preclude continued employment in the same job. If there is sufficiently low expected marginal utility of consumption conditional on such a move, a positive transferable pension is called for, even though this decreases the insurance available against low productivity with the first-period employer.

Second, we examine the case where individuals only leave an employer for a better offer. Once we recognize that wages elsewhere can exceed marginal productivity with the first-period employer at the same time that both wage and productivity are low, there is a reason for the optimal contract to contain a pension payment for a worker who leaves his first-period employer. The extent of these payments depends on the observability of both productivity with the first-period employer and the wage offer elsewhere. With both of these observable, the optimal contract implies that the worker moves if and only if his wage elsewhere exceeds his marginal productivity with the original firm. The transferable pension is positive when the wage elsewhere is sufficiently low. This same outcome is achievable with suitable counteroffers by the firm when the wage elsewhere is observable but productivity with the first-period employer is not observable.

The analysis becomes more complicated when there are limits on the observability of the outside wage offer. For reasons of tractability, we only consider cases where productivity with the first-period employer is observable. Two cases are analyzed—where the outside offer is never observable by the firm and where it is not observable in time to make a counteroffer but becomes observable in time to condition the transferable pension on earnings elsewhere.

When the transferable pension can depend on earnings elsewhere, low earnings and low productivity result in constant consumption over time, a positive transferable pension, and mobility if and only if wages elsewhere exceed productivity with the first-period employer. With low productivity and a high outside offer, there is mobility and no transferable pension. With high productivity, there is no transferable pension and a wage offer from the first-period employer below productivity. Mobility depends on the comparison of wage offers. Thus, workers sometimes move when wages elsewhere are below productivity with the first-period firm.

When the outside offer is totally unobservable, both types of inefficiency become possible. When productivity is high, the worker will sometimes move to a wage below his productivity. When productivity is low, the worker sometimes will stay with his first-period employer even though the wage exceeds productivity. There will be a positive transferable pension for some of his productivities.

Notes

1. To see the effects of private savings on optimal social insurance, contrast our papers (1978) and (1982).

2. This is not quite right if deferred payments are made only in period 3, but there is no reason in the model for such a restriction.

3. We are indebted to Robert Merton for pointing out this aspect of this example.

4. For analyses of mobility with alternative contractual assumptions, see Mortenson (1978), Diamond and Maskin (1979), Hall and Lazear (1981), Moore (1982), Deere (1983), and Spinnewyn (1982).

5. We also assume that f has a sufficiently large support to allow all the cases we consider below. In particular, we assume that $m_2^{\prime}$ might be larger or smaller than m_2 for all values of m_2 and m_2 might be larger or smaller than $2w_1$ for the equilibrium value of w_1.

6. Hall (1982) has estimated that approximately half of current jobs will have lasted for at least 20 years by their termination. From his numbers, one can also infer the obverse, that approximately half of workers will never hold any single job for as long as 20 years.

7. Bulow and Scholes (1984) consider a model where implicit contracts between firm and worker are replaced by implicit contracts among workers. Because of the complications of union institutional structure and politics we confine our discussion to nonunion plans. However, it is tempting to speculate on the large differences between single-employer and multi-employer plans.

8. Firing a salesman to prevent his collecting commissions on a previously negotiated deal did lead to damages in Fortune v. National Cash Register Co., Supreme Judicial Court of Massachusetts 1977, 373 Mass. 96, 364 N.E. 2d 1851. We are indebted to Melvin Eisenberg for this reference.

9. For models of reputation for consumer good quality with a similar character, see Dybvig and Spatt (1980), Klein and Leffler (1981), and Shapiro (1981).

10. This appearance is clearest in the use of profit sharing for defined contribution plans.

11. The ERISA rules to limit backloading are an example of such consumer protection. It is curious that the rules assume no wage growth.

Comment Robert C. Merton

1. Introduction

In a world of full information and perfect markets, where all assets (including human capital) are freely tradable, private pensions provide nothing more than another way for individuals to save.[1] With a full complement of risk-sharing securities available, the worker can fully offset or modify any particular form of payouts prescribed by the pension plan. Hence, the type of pension plan offered would be a matter of indifference to workers. Like the Modigliani-Miller theorem for corporate liabilities, the optimal choice of pension plan would at most be a function of the tax laws and perhaps certain kinds of transactions costs. In such an environment and in the absence of explicit contracts to the contrary, the spot-market view of total employee compensation would obtain where this period's wages plus incremental vested retirement benefits are equal to the worker's current marginal product.[2] For pension plans to have a greater functional significance than that of "just one more security," there must

Robert C. Merton is J. C. Penney Professor of Management at the Sloan School of Management, Massachusetts Institute of Technology, and research associate, National Bureau of Economic Research.

1. The saving component is obvious for a defined contribution plan. It also takes place for a defined benefit plan whether or not the firm funds the plan with a pension fund since the (pension) liability issued to the worker in lieu of other cash compensation provides resources to the firm for financing investment just like the issuing of any other liability.

2. This assumes that firms are competitors in the labor market. For a development of the spot market theory with respect to pension liabilities, see Bulow (1982).

be important market imperfections, and the most likely place for these imperfections to occur is in the labor market.[3]

There are, of course, severe impediments to the trading of human capital. The two most prominent explanations for this nontradability are (1) the moral hazard or incentive problem that having once sold off the rights to their earnings, workers will no longer have the incentive to work, and (2) the broad social and legal prohibition of (indentured) slavery whether voluntary or otherwise. The well-known effects of this nontradability are to "force" workers to save more than they might otherwise choose and to cause them to bear much of the risk (both systematic and nonsystematic)[4] of their human capital. In addition to distorting the consumption-saving choice, the nonoptimal risk bearing of the risks of human capital may cause inefficient investment of resources in developing human capital.[5]

In "Insurance Aspects of Pensions," Diamond and Mirrlees explore the possible role of private pensions in insuring the worker against some of the risk associated with his nonmarketable human capital.[6] Their analysis leans heavily on the Harris and Holmstrom (1982) theory of implicit contracts in the labor market. By assuming away the incentive problems associated with the effective sale of human capital, Diamond and Mirrlees focus on contracts that taken into account the limitations on workers to bind themselves now to work in the future for less than they could otherwise earn in the absence of such contracts.

As Diamond and Mirrlees themselves note, their model uses a number of assumptions that are wholly unrealistic even by the standards of casual empiricism. Many of these assumptions are constructive in that they merely simplify the analysis without severely distorting the validity of their central conclusions. Others, however, are crucial in reaching their results, and it is on these that I focus this discussion.

As will be shown, the structure of the labor contracts derived by Diamond and Mirrlees (and, for that matter, by Harris and Holmstrom) are isomorphic to various put and call option contracts. Reformulating

3. If the state is paternalistic and wants to avoid the free rider problem, it may choose to provide tax incentives to induce saving through pensions together with nonassignment of the pension and penalties for early withdrawal to enforce the availability of adequate unencumbered assets for the individual to fund his retirement. Even with no labor market imperfections, these distortions could lead to optimal characteristics for pension plans that minimize the effects of the distortions.

4. Nonsystematic risk is risk that can be eliminated through diversification when there are available adequate risk-sharing financial instruments. Systematic risk is risk that cannot be eliminated by these means and hence is a risk that must be borne by the economy even with perfect insurance markets.

5. Thus, a worker may spend real resources to reduce the risk of his human capital by pursuing training that makes him able to undertake a wider variety of jobs. Such diversification of the worker's human capital would not be an optimal allocation if it simply reduces nonsystematic risk.

6. See Nalebuff and Zeckhauser (1981), especially App. 1, for a similar analysis.

their findings in this context will help to shed light on the sensitivity of their conclusions to certain of their assumptions. Moreover, the extensive literature on the evaluation of options permits the derivation of comparative statics results that might otherwise not be apparent.[7] The options analogy is developed first in section 2 within the context of the "mobility-threat" model where the equilibrium contract leads to no actual changes in employer by the worker. In section 3, the analogy is extended to the more complex cases of exogenous and endogenous mobility where workers will change employers under the appropriate equilibrium conditions. Section 4 provides a brief summary and overview of the Diamond-Mirrlees model.

2. Optimal Labor Contracts Viewed as Options: Mobility Threats

The model presented in section 11.2 of the Diamond-Mirrlees paper assumes a three-period life for each worker (two work periods and a retirement period) and that workers are risk averse with an additive separable and symmetric utility function for lifetime consumption. It is assumed that firms have full access to a well-functioning capital market and that all securities are priced to yield the same expected return. This "common" expected return is assumed to be zero in real terms, and all contracts (explicit or implicit) are expressed in real terms. In sharp contrast, workers are assumed to have no access to the capital market and own no assets other than their human capital. Thus, workers cannot borrow, and they can only save if the firm (acting as an intermediary) does it for them. In this version of the model it is assumed that a worker's marginal product in period t ($t = 1, 2$) is the same for all firms (i.e., $m_t = m_t'$). (This assumption is relaxed in later sections.) Firms (other than perhaps the worker's current employer) are always willing to pay the worker his marginal product (to them), m_t', if the employee is willing to move.

To locate and understand the model, consider first the case where there are no restrictions on contracts between the worker and the firm. In this case, the only constraint on the contract is that its present value, PV, satisfy

$$PV = \mathrm{m}_1 + E(m_2), \tag{1}$$

which is the value to the firm at time 1 of receiving the labor of the worker throughout his work life. Because there is no risk premium paid for risk bearing and workers are risk averse, the optimal (worker utility maximizing) contract provides for the worker to bear no risk. By the assumptions of symmetric utility and a zero interest rate, the optimal contract would pay the worker a first-period wage, $w_1 = PV/3$; a guaranteed second-period wage, $w_2 = PV/3$; and a pension benefit payment, $b = PV/3$. In return, the worker would agree to work for his original employer for his

7. See Smith (1976) and Mason and Merton (1984) for surveys of the option literature including its application in the evaluation of nonfinancial market options.

entire work life, independently of what his (currently unknown) marginal product, m_2, turns out to be in period 2. If the worker were permitted to enter into such a binding agreement with his employer, then this unrestricted optimal contract is feasible.

The worker cannot, however, bind himself to work for his current employer in the second period. The employer must, therefore, take into account that the worker will leave in the event that the total compensation offered by a competing firm in the second period plus any transferable component of his pension from his original employer, $m_2' + b'$, exceeds the total compensation 2PV/3 provided by the contract with his original employer. Thus, unless the probability that $m_2' + b' > 2PV/3$ is zero, the unrestricted optimal contract is not feasible and the worker must bear some risk.

Given this constraint and the condition that $b' \geq 0$, Diamond and Mirrlees (eq. [2]) show that the optimal contract calls for a (combined retirement and) second-period compensation given by

$$W(m_2) = \max(2w_1, m_2) \tag{2}$$

and a transferable pension $b' = 0$.

A call option on a security (or payment) gives its owner the right to buy the security at a specified price I (the "exercise price") as of a given date. The right to buy also implies the right *not* to buy. If the owner chooses not to buy, the call option expires, worthless. Thus, the payoff to a call option on its expiration date is given by $\max(0, X - I)$, where X is the price of the security on that date. Rewriting (4) as

$$W(m_2) = 2w_1 + \max(0, m_2 - 2w_1), \tag{2'}$$

we have that the optimal contract provides for a guaranteed payment of $2w_1$ plus a call option on the worker's second-period marginal product with an exercise price of $2w_1$. If $C(X, I)$ denotes the market value of a call option on a security with current value X and exercise price I,[8] then the budget constraint (1) requires that the value of what the worker receives—first-period wage, w_1, plus second-period guaranteed payment, $2w_1$, plus the call option $C[E(m_2), 2w_1]$—equals the present value of his human capital. That is,

$$3w_1 + C[E(m_2), 2w_1] = PV = m_1 + E(m_2). \tag{3}$$

Because the payoff to a call option is nonnegative, $C[E(m_2), 2w_1] \geq 0$ with equality holding if and only if $\mathrm{pr}\{m_2 > 2w_1\} = 0$. It follows, therefore, that $w_1 < PV/3$ unless $\mathrm{pr}\{m_2 > 2[m_1 + E(m_2)]/3\} = 0$.

8. In the special case assumed by Diamond and Mirrlees of risk-neutral security pricing and a zero interest rate, $X = E(\tilde{X})$ where $\tilde{X}$ is the random variable value of the security at the expiration date of the option and $C(X, I) = E[\max(0, \tilde{X} - I)]$.

Armed with this description, the following comparative statics results follow directly from budget constraint (3): If we hold fixed the distribution of $m_2 - E(m_2)$, then by the implicit function theorem, we have that

$$\frac{dw_1}{dE(m_2)} = \left[1 - \frac{\partial C}{\partial X}\right] / \left[3 + 2\frac{\partial C}{\partial I}\right]. \tag{4}$$

As is well known for call option valuations, $0 \leq \partial C/\partial X \leq 1$ and $0 \geq \partial C/\partial I \geq -1$. Hence, $dw_1/dE(m_2) \geq 0$ with strict equality holding only if $\text{pr}\{m_2 < 2w_1\} = 0$. Thus, ceteris paribus, a higher expected second-period marginal product leads to a higher current wage and a higher second-period "guaranteed" wage floor.

Consider instead a "mean-preserving" spread on m_2, where $E(m_2)$ is held fixed and the uncertainty about the second-period wage, σ, increases. From (3),

$$\frac{dw_1}{d\sigma} = -\frac{\partial C}{\partial \sigma} / \left(3 + 2\frac{\partial C}{\partial I}\right). \tag{5}$$

Again a well-known result for call options is that $\partial C/\partial \sigma \geq 0$ with strict equality holding only if $\text{pr}\{m_2 > 2w_1\} = 1$ or 0. Thus, the greater the uncertainty about the worker's second-period marginal product, the smaller is the current wage and the lower is the second-period wage floor. It follows immediately that the amount of risk borne by the worker increases the greater the uncertainty about his future marginal product.

Although pursued no further here, the known relations for option prices can be used to derive other comparative statics results. It should be noted that the applied properties for option prices do not depend on the assumptions of risk-neutral pricing of securities in the capital markets or a zero interest rate. Hence, relaxing these assumptions will not affect the basic comparative statics results, although the existence of a positive risk-return trade-off may alter somewhat the structure of the optimal contract.

A put option gives its owner the right to sell a security at a specified exercise price on a specified date. The payoff to a put option on its expiration date is given by max $(0, I - X)$, where I is the exercise price and X is the price of the security on that date. The standard functional description of a put option is that of insurance because the purchase of a put protects the owner of the underlying security against losses (in value) below the specified floor I. To see this insurance feature of the optimal contract, I rewrite (2) as

$$W(m_2) = m_2 + \max(0, 2w_1 - m_2). \tag{2''}$$

Hence, the optimal contract provides the worker with a put option or insurance on the future wage he would otherwise earn if the labor market were simply a spot market. Of course, the worker pays for this insurance out of his current wage. That is, the budget constraint (1) can be written as

(6) $$w_1 = m_1 - P[E(m_2), 2w_1],$$

where $P(X, I)$ is the value of a put option with an exercise price of I on a security with current value X. Although (6) follows directly from (3) by the parity theorem for option prices,[9] expressing the budget constraint in the form of (6) makes more apparent the result that the first-period wage w_1 must always be less than the first-period marginal product m_1.

This result underscores how binding the assumed constraint of no borrowing by workers (embedded in the model by requiring $b' \geq 0$) can be in the determination of the optimal feasible contract. Thus, at the extreme where the worker's current marginal product $m_1 = 0$, the worker's current wage $w_1 = 0$, and moreover he can obtain no insurance for his future earnings. This is so no matter how large is the worker's future expected productivity $E(m_2)$. In a richer model where the worker can influence his current and future marginal products through training or choice of career path, it is readily apparent that the no-borrowing constraint (which is central to the model here) will cause severe distortions of the optimal labor force configuration. That is, with this constraint, workers achieve substantial benefits in terms of both the level of early life consumption *and* the reduction of risk surrounding later life consumption by choosing a job pattern with relatively high early life marginal product m_1 even at the expense of a large reduction in expected future marginal product, $E(m_2)$.

3. Optimal Labor Contracts Viewed as Options: Exogenous and Endogenous Mobility

In the Harris-Holmstrom type model where the worker's marginal product is the same for all firms, the worker will never change employers in equilibrium and transferable pensions play no role in intermediating the risk of human capital. However, if the worker must change employers for exogenous reasons, or if the worker can have different marginal products at different firms, then the worker may move in equilibrium. Although Diamond and Mirrlees discuss the cases of exogenous and endogenous mobility separately, the discussion here combines these cases in a single model.

As with the mobility-threat analysis, Diamond and Mirrlees derive the optimal contract under the condition that the firm can offer a transferable pension b' if the employee moves and a total compensation floor if the employee stays. If the worker is forced to move ("exogenous mobility"), then he receives $m' + b'$. The probability of such a move, p, is given exogenously. Otherwise, the worker will move only if his total compensation at the new firm, $m_2' + b'$, exceeds that offered by his original employer.

9. See Smith (1976) or Mason and Merton (1984) for proof of the theorem by arbitrage. In the special case of a zero interest rate, the theorem requires that $C(X, I) = X - I + P(X, I)$.

Although not stated explicitly, Diamond and Mirrlees make an important assumption about the structure of the labor market. In the initial period, all firms compete equally to hire the services of the (unattached) worker. In the second-period compensation negotiations, Diamond and Mirrlees assume an asymmetry between the worker's current employer and all other potential employers which gives the worker's current employer an advantage. They postulate those firms which attempt to hire a worker away from his original firm must pay a wage equal to his full marginal product m_2' whereas the current employer need only match the total compensation provided to the worker for moving, $m_2' + b'$, to induce the worker to stay. Since the transferable pension (if any) b' is a "sunk" cost which the firm will have to pay if the employee moves, the current employer derives a benefit from this monopsony power equal to $(m_2 - m_2')$, which is nonnegative whenever it is optimal for the worker to remain with his original employer.

As in the discussion of the mobility threat case in the previous section, I begin with an analysis of the worker's situation when there are no restrictions on contracts. In the absence of monopsony power, the present value of the worker's human capital is given by

$$W_1 = m' + pE(m') + (1 - p)E[\max(m_2, m_2')]. \tag{7}$$

As before, the unrestricted optimal contract is riskless to the worker with $w_1 = W_1/3$, $w_2 = W_1/3$, and $b = b' = W_1/3$; and in return the original employer receives the worker's marginal product whether he moves or not. Introduction of the Diamond-Mirrlees asymmetry assumption might seem to lower the present value of the worker's human capital to equal $m_1 + pE(m') + (1 - p)E(m_2')$. However, because firms are competitive in the initial period, they must pay for this right to "exploit" the worker in the second period. That is, they will pay the worker a "signing bonus" given by

$$(1 - p)E[\max(m_2, m_2') - m_2'] = (1 - p)E[\max(0, m_2 - m_2')]. \tag{8}$$

By inspection of (8), $\max(0, m_2 - m_2']$ is the payoff to a call option on a payment of $m_2 - m_2'$ with a zero exercise price, and we denote its price by $C[E(m_2 - m_2'), 0]$. (Note: Unlike the "usual case" of a call option on a limited liability instrument, $C[E(m_2 - m_2'), 0] > \max[0, E(m_2 - m_2')]$ if the $\mathrm{pr}\{m_2 > m_2'\} > 0$. Hence, even if the expected future marginal product of the worker with his current employer, $E(m_2)$, is less than or equal to his expected future marginal product elsewhere, $E(m_2')$, the signing bonus will be positive.) In the Harris-Holmstrom case, where $m_2 = m_2'$, there is no value to the firm of obtaining this second-period monopsony power, and hence its assumption has no influence on the analysis of the mobility threat case.

As Diamond and Mirrlees show, when the worker's second-period earnings, y (whether with the current employer or not), can be observed ex post at the end of the second period, the optimal labor contract calls for a transferable pension given by $b'(y) = \max(0, 2w_1 - y)$ if the worker moves and a compensation package equal to $\max(2w_1, m_2)$ if the worker stays with his original employer. If the worker is not forced to move, then this contract leads to the worker's staying with his current employer if and only if $m_2 \geq m_2'$.

The worker's second- (and retirement) period compensation is given by

$$W(y) = y + \max(0, 2w_1 - y), \tag{9}$$

which in terms of structure is the same as the Harris-Holmstrom case (2″). However, in (9), $y = m'$ if the worker is forced to move and $y = m_2'$ otherwise. From (9), it would appear that, from the worker's perspective, his second-period compensation does not depend on his second-period marginal product with his original employer, m_2. This is true in the ex post sense of the uncertainties surrounding y, given the wage floor, $2w_1$. However, the ex ante distributional characteristics of m_2 do affect $W(y)$ because they affect the first-period wage and hence the second-period wage floor.

To determine the first-period wage w_1, equate the value that the firm receives in return for the contract (i.e., the worker's first-period marginal product plus the right to exploit the worker in the second period) to the cost of the contract (i.e., the worker's first-period wage plus the insurance provided by the transferable pension and the wage floor). Expressed in terms of the option value equivalent of the contract, we have that

$$\begin{aligned} w_1 &= m_1 + (1 - p)C[E(m_2 - m_2'), 0] \\ &\quad - pP[E(m'), 2w_1] - (1 - p)P[E(m_2'), 2w_1]. \end{aligned} \tag{10}$$

By inspection, the distributional characteristics of m_2 (and not just its expected value) affect the first-period wage through their influence on the call option price, which reflects the value of the signing bonus.

Comparative statics results along the lines of the previous section can be computed from (10). I do so here only with respect to the effect of a change in p as the means of providing a few comments on the Diamond-Mirrlees exogenous mobility model presented in their section 11.3. From (10), we have that

$$\begin{aligned} \text{sign}\left(\frac{dw_1}{dp}\right) &= \text{sign}\{P[E(m_2'), 2w_1] \\ &\quad - P[E(m'), 2w_1] - C[E(m_2 - m_2'), 0]\}. \end{aligned} \tag{11}$$

In the case of exogenous mobility alone (i.e., $m_2 = m_2'$), the sign of dw_1/dp depends on the value of a put option on the worker's marginal product when he is forced to move, m', relative to the value of a put option of identical terms on the worker's marginal product when he can either stay or move, m_2'. Although it is, of course, conceivable for m' to have a more favorable distribution than m_2 or m_2', it would appear to be a rather strained definition of a forced or exogenous move if the worker receives a wage m' in excess of what he would have earned otherwise. This belief is further reinforced by the model's assumption that among all the firms that might offer the worker a job in the nonforced case, the best offer will be a wage equal to the one available from the worker's current employer. Moreover, if, as Diamond and Mirrlees at one point assume, the exogenous move is the result of a disability, then the case for $m' < m_2$ would seem especially compelling.

If one postulates that $m' < m_2'$, then, the value of the put option insurance on m' is greater than the value of the corresponding insurance on m_2'. From (11), we have unambiguously that $dw_1/dp < 0$. That is, if the probability of ending up in the disadvantaged state of being forced to leave your current employer increases, then the worker must (and is willing to) pay more to insure against the lost income in this state by accepting both a lower first-period wage and a lower floor on guaranteed compensation.

Making this incremental assumption also provides some insight into Diamond and Mirrlees's proposition 1, which provides sufficient conditions for the optimal transferable pension, b', to be positive. By assuming that $m_2 < [m_1 + (1 - p)E(m_2)]/3/2 - p)$ for all m_2, they ensure that $m_2 < 2w_1$, the equilibrium wage floor. Hence, in the event of no forced move, the put option on m_2 will always be exercised and the worker receives a guaranteed second-period wage with no uncertainty. If the transferable pension is restricted to be a constant, then the budget constraint is

$$w_1 = m_1 - pb' - (1 - p)[2w_1 - E(m_2)] \tag{12}$$

because $P[E(m_2), 2w_1] = 2w_1 - E(m_2)$ in this case. From (12), if $b' = 0$, then $2w_1 = [m_1 + (1 - p)E(m_2)]/[3/2 - p]$, which, by hypothesis, strictly exceeds the maximum possible marginal product to be earned in period 2. Thus, it pays to transfer at least the residual value to a transferable pension, which reduces the risk of lost income when forced to move.

By dispensing with the requirement that $b' > 0$ be a constant and positive for all possible m' and assuming that $m' < m_2$, the role for a transferable pension is established without the extreme conditions of proposition 1.

Although Diamond and Mirrlees examine a variety of other cases where various marginal products are or are not observable, I have focused exclusively on the case where the transferable pension can be made a function of the ex post second-period earnings of the worker and the wage floor provided to the worker, if he stays, is a constant. This choice was not arbitrary. This case surely establishes a nontrivial role for private pensions as a means of reducing risk to workers. The derived rules for the transferable pension and wage floor are simple and yet appear to be reasonably robust.

Thus, while the level of the floor and maximum transferable pension depend on the symmetry of the worker's utility function to be a true optimum, they do not depend on the specific form of the worker's utility function, which some of their other derived rules do. Having the transferable pension benefit depend on ex post earnings seems to be a practical possibility. By inspection of income tax returns, the pension-paying firm could verify the later year earnings for computing pension benefits. Although there is in principle an incentive for the worker to cheat by hiding his income, this possibility would not appear to be serious because it is difficult for workers to avoid declaring wages or benefits reported on W-2 forms. Moreover, to do so would require that the worker cheat on his federal income taxes and thereby expose himself to those penalties as well. The derived plan has the further virtue of not requiring the pension-paying firm to distinguish between involuntary moves (e.g., disability) and voluntary moves.

4. Summary and Conclusion

The Diamond and Mirrlees analysis should stimulate much-needed additional research into the role of pensions when there are significant labor market imperfections. Although their model is almost too simple, many of their assumptions can be relaxed without seriously affecting their basic conclusions. I applaud their attempt to bring institutional and legal constraints into the discussion of feasible contracts for their model.

Perhaps with a bit of irony, such real world legal constraints may rule out their assumption that the firm can "exploit" its current employees by paying them $m' + b'$ when $m_2 > m_2'$. If, for example, $m_2 > m_2' > 2w_1$, then it is likely that the firm will be hiring new workers in addition to their current labor force. Since by their other assumption firms must pay new workers their full marginal product, it is not clear that it would stand a legal test for the firm to pay "new" workers (who are otherwise the same) more than "old" workers. The sensitivity of their contract schemes to this issue warrants further study.

Diamond and Mirrlees raise the important issue of the firm's defaulting on its labor contracts. The manifest impact of such a default is that the workers may not be paid the transferable pension and "wage floor" benefits promised in both the explicit and implicit parts of the labor contract. If workers fully recognize this possibility ex ante, when the contract is negotiated, they will receive compensation for this risk in the form of either a higher first-period wage or larger promised future benefits. Even in this case, however, it is straightforward to show that the workers' expected utility will fall relative to the case where such defaults are ruled out. If, as would seem reasonable, the fortunes of the firm and the future marginal product of its workers are strongly positively correlated, it is more likely that default will occur in precisely those states where the worker most needs the promised benefits. Thus, the utility loss from the possibility of default could be substantial.

There is another—perhaps latent—impact of default that may significantly limit the magnitude and duration of the future promised benefits component of labor contracts. Bankruptcy is, of course, bad for the firm—but it is less bad than if the owners (shareholders) were required to contribute additional sums to the firm in order that it could fulfill its obligation and avoid bankruptcy. As has been shown in the finance literature, it is not the shareholders who lose in a bankruptcy, but the other liabilityholders who are not paid what they are promised. From this follows the well-known result that if the riskiness of the firm's assets increases, then there will be a transfer of value from the current nonequity liabilityholders to the stockholders, and this occurs even if such an increase has no effect on the overall market value of the firm. Thus, the managers of limited liability firms have an incentive to increase the riskiness of their assets (even if there is no compensating higher expected return on the assets).[10] In the Diamond-Mirrlees context, the workers are explicit nonequity liabilityholders of the firm with respect to their vested pension benefits and implicit liabilityholders with respect to the firm's promise of a floor on future wages. Hence, they face a potential moral hazard problem with respect to the firm not unlike the one that rules out or severely limits the sale of their own human capital.

Since the potential for gain to the stockholders from such a "liability-induced" shift in the risk of the firm's assets is an increasing function of both the size and the maturity of the nonequity liabilities, the moral hazard problem can be reduced by limiting the magnitude of the liabilities and making them of relatively short duration. For explicit corporate liabilities, the problem is further reduced by the introduction of indenture restrictions that limit the types of investments the firm can make without the liabilityholders' approval. A relevant example is a corporate pension plan that requires that a specified portion of the firm's assets be segregated in a pension fund; gives the pension liabilityholders first claim on these assets; and sets guidelines for the types of assets held in the fund. Thus, with the possibility of default, optimal labor contracts will have a smaller proportion of total compensation in the form of promised future benefits than would be predicted by the Diamond-Mirrlees analysis. Moreover, the relative proportion of explicit to implicit contract benefits will also be larger, which suggests that vested and transferable pensions may have an even more important role in improving the risk-bearing opportunities for workers.

10. Thus, even if all securities and assets are priced in a risk-neutral fashion, the presence of bankruptcy possibililties may induce what appears to be "risk-loving-like" behavior on the part of managers in their selection of the firm's assets.

If the moral hazard problem surrounding default by the firm can be neglected, then the manifest impact of bankruptcy can be formally integrated into the current Diamond-Mirrlees model by recognizing that the workers, as part of the implicit labor contract, grant a put option to the firm with an exercise price equal to the total promised second-period benefits.

In the case of default on pension benefits, which are a legal liability of the firm, the firm would face actual bankruptcy and the underlying security for the put option would be the firm's assets. In the (perhaps more interesting) case of default on implicitly contracted wage floors, there would be no formal bankruptcy. While it is not always clear what the underlying security for the put is in this case, I would agree with Diamond and Mirrlees that it is probably the value of the intangible asset called "reputation" lost by the firm when it does not meet its implicit contract obligations. Such an analysis might also explain why some firms when facing hard times choose to employ all their long-term workers part time instead of firing some workers and keeping others full time. In effect, like bondholders in a single debt issue, partial employment permits all workers with defaulted implicit claims to share what amount is available, whereas full employment of some and no employment for others would be the equivalent in a regular bankruptcy of randomly paying some bondholders full face value and others nothing at all. Using the established results from option pricing, one could perhaps identify the type of firms (e.g., by risk characteristics) in which implicit labor contracts with floors are likely to be important.

Finally, I would note that the derived payoff structure to the transferable pension benefit, $b' = \max(0, 2w_1 - y)$, looks remarkably similar to the structure of a defined benefit private pension plan that is integrated with social security.[11] In effect, the Diamond-Mirrlees pension benefit is equivalent to a defined benefit plan integrated with other private plans instead of social security. Since they derive this structure as the solution of an optimal plan, it may be possible to derive similar normative properties, heretofore unrecognized, for integrated pension plans.

References

Bulow, J. I. 1982. What are corporate pension liabilities? *Quarterly Journal of Economics* 97:435–52.

Bulow, J. I., and Scholes, M. S. 1984. Who owns the assets in a defined benefit pension plan? In *Financial aspects of the United States pension*

11. See Merton et al. (1984) for a discussion of the insurance aspects of integrated pension plans and their isomorphic relation to put options.

system, ed. Z. Bodie and J. B. Shoven. Chicago: University of Chicago Press (for NBER).

Bulow, J. I.; Scholes, M. S.; and Menell, P. 1984. Economic implications of ERISA. In *Financial aspects of the United States pension system,* ed. Z. Bodie and J. B. Shoven. Chicago: University of Chicago Press (for NBER).

Burton, S. 1980. Breach of contract and the common law duty to perform in good faith. *Harvard Law Review* 94:369–405.

Deere, D. 1983. Labor market distortions, labor turnover, and the role of unemployment insurance. Ph.D. dissertation, Massachusetts Institute of Technology.

Diamond, P. and Maskin, E. 1979. An equilibrium analysis of search and breach of contract. *Bell Journal of Economics* 10:212–316.

Diamond, P., and Mirrlees, J. 1978. Social insurance with variable retirement. *Journal of Public Economics.*

______. 1982. Social insurance with variable retirement and private savings. MIT Working Paper.

Dybvig, P., and Spatt, C. 1980. Does it pay to maintain a reputation? Unpublished.

Hall, R. 1982. The importance of lifetime jobs in the U.S. economy. *American Economic Review* 72:716–24.

Hall, R., and Lazear, E. 1984. The excess sensitivity of layoffs and quits to demand. *Journal of Labor Economics* 2:000–000.

Harris, M., and Holmstrom, B. 1982. A theory of wage dynamics. *Review of Economic Studies* 49:315–33.

Hart, O. 1983. Optimal labour contracts under asymmetric information: An introduction. *Review of Economic Studies* 50:3–35.

Klein, B., and Leffler, K. 1981. The role of market forces in assuming contractual performance. *Journal of Political Economy* 89:615–41.

Mason, S., and Merton, R. C. 1984. The role of contingent claims analysis in corporate finance. Harvard Business School Working Paper (February).

Merton, R. C.; Bodie, Z.; and Marcus, A. J. 1984. Pension plan integration as insurance against social security risk. NBER Working Paper 1370.

Mirrlees, J. 1982. Migration and optimal income taxes. *Journal of Public Economics* 18:319–41.

Moore, J. 1982. Optimal labor contracts when workers have a variety of privately observed reservation wages. Birkbeck College Discussion Paper.

Mortenson, D. 1978. Specific capital and labor turnover. *Bell Journal of Economics* 9:572–86.

Nalebuff, B., and Zeckhauser, R. 1981. Involuntary unemployment reconsidered: Second-best contracting with heterogenous firms and workers. Mimeographed. Cambridge: Harvard University, Department of Economics, October.

Shapiro, C. 1983. Premiums for high quality products as rents to reputation. *Quarterly Journal of Economics* 97:659–79.

Smith, C. W. 1976. Option pricing, a review. *Journal of Financial Economics* 3:3–51.

Spinnewyn, F. 1983. Long-term labor contracts viewed as saving workers within the firm: A dual approach. CORE Discussion Paper 8308.