10 Pensions and the Retirement Decision

Barry Nalebuff
Richard J. Zeckhauser

10.1 Introduction

Pensions facilitate labor contracting. They provide an additional instrument beyond wages for attracting, sorting, and motivating workers. The key difference from other forms of labor compensation is that pensions are paid during the last years of one's life, usually as a contingent claim with payments continuing as long as one lives. The late-payment feature has the advantage of allowing an individual's reward to depend not only on the present period but also on future experiences. Its disadvantage is that it may hinder a worker's lifetime allocation of income unless he can trade on well-functioning capital markets.

The contingent claims feature has obvious risk-spreading advantages. However, contingent claims markets are often flawed in that they change incentives for individuals to engage in various types of behavior and induce them to "purchase" inappropriate claims. The problem of moral hazard is not severe here. Individuals have quite adequate incentives, apart from the pensions they will receive, to increase their survival. Similarly, problems of adverse selection may be limited, because pensions tend to be universal in a workplace, and pension considerations are unlikely to be the critical factor in job choices. Moreover, individuals are not likely to have substantial information about their life expectancy early in life, at the time pension benefits start to be accrued.

Barry Nalebuff is a junior fellow of the Society of Fellows of Harvard University. Richard J. Zeckhauser is professor of political economy at Harvard University's John F. Kennedy School of Government and a research associate of the National Bureau of Economic Research.

We received great insight and support from Andrew Nevin. Brian Sullivan provided excellent research assistance. John Dunlop, Jerry Hausman, Larry Summers, and seminar participants at Princeton, UCSD, and NBER provided stimulating comments. Cathy Lydon worked magic with the word processor, Nancy Jackson with the red pencil.

Given their contingent claims nature, and the fact that they are paid at the end of one's life, the most direct labor market effect of pensions may be on individuals' retirement decisions. In most pension plans, the per period benefits and expected payouts that a worker receives depend significantly on the age at which he retires. The central purpose of this analysis is to explore the effects of pensions on retirement decisions and to discuss the implications of those effects for policy choice.

A roadmap to our paper may be of assistance. Section 10.1.1 provides a brief description of our major results. Section 10.1.2 explores seven factors influencing individuals' retirement decisions, the factors that motivate our subsequent formulations. We then provide a capsule overview of the historical reasons leading to the introduction of pensions. The section concludes with a summary of the labor market effects apart from retirement decisions.

Sections 10.2, 10.3, and 10.4 present our substantive results. Each part is self-contained. All results are motivated by intuition. When proofs seem complicated, look for the intuitive arguments nuzzling with the mathematics.

Section 10.2 examines the effects of pensions on retirement. Pensions are first viewed as forced savings. As such, pensions encourage retirement; indeed, that is one of their primary purposes. Since it is difficult to monitor a worker's true disutility or work or to make contracts in which the worker commits himself to retire, pensions are used to help induce appropriate retirement behavior. How does the value of "pension savings" vary with age? Section 10.2.3 begins this inquiry, taking typical defined benefit pension plans—as revealed in a survey described in the appendix—and examining the relationship they produce between retirement age and expected benefits. Since individuals who choose to retire at different ages may have different (possibly unobservable) characteristics (e.g., how long they expect to live), we emphasize the distinction between the actuarial treatment of individuals and the actuarial treatment of retirement cohorts.

Section 10.3 focuses on populations that are heterogeneous with respect to such factors as preferences or life expectancy. It explores the design of optimal pensions (e.g., the way annual benefits should vary with retirement age), given heterogeneity. If there are asymmetries of information, a second-best outcome in which allocative efficiency tugs against full risk spreading must be expected. Again, confusion reigns on the issue of actuarial fairness. We hope to dethrone it just a bit.

Section 10.4 examines the effects of requiring pension offers to be the same among workers with dissimilar preferences, life expectancies, and productivity profiles. Social security programs and numerous regulations now in effect in the United States encourage or require common structures. We show how all workers may lose from such commonality, no matter what plan is adopted.

Section 10.5 presents the conclusions.

10.1.1 Summary of Results

Our paper includes several models. We list here some of the major results, grouped under three headings.

Appropriate Retirement (10.2.2)

1. It is a stylized fact that wages frequently exceed productivity in the later periods of work life. Because workers are unable to commit themselves to retire under appropriate circumstances (i.e., when disutility exceeds productivity, not wages), they will choose to work too long. Pensions can be used to force workers to save more than they want to save. Since these excess savings are accessible only on retirement, workers choose to retire earlier. In the new equilibrium, firms can pay higher wages, raising the worker's lifetime expected utility.

Optimal Pensions with Unmonitorable Information
(10.2.3, 10.3.1, 10.3.2)

2. *Actuarial treatment of cohorts.* Most defined benefit pension plans appear to be actuarially unfair to late retirees. However, when workers can estimate their life span, those who expect to live longer choose to retire later and the pattern of actuarial benefit may be reversed.

3. *Actuarial treatment of individuals.* The structure of optimal pension plans, that is, those that maximize ex ante expected utility, must make it actuarially unfavorable for an individual to retire later.

4. *Disutility unmonitorable.* The optimal pension plan when disutility of work is unmonitorable offers a benefit that rises with retirement age up to a point and is then level. This point, in effect, is the maximum retirement age.

The rising portion of the pension curve sacrifices risk spreading to discourage workers from retiring too early. Under the optimal plan, assuming that death dates do not correlate with disutility, cohorts that retire later receive lower expected pension benefits.

Common Pension Plan for Heterogeneous Populations (10.4)

5. Consider the optimal pension plans for each group of workers with different characteristics. If, as required and encouraged by law, these plans are merged, then problems of adverse selection and moral hazard may make *all* groups worse off. Indeed there may be no common pension plan that is superior for any group of workers to what they each received when treated separately.

10.1.2 Why People Retire

The purpose of this analysis is to identify the role of pensions in affecting individuals' retirement decisions. At the outset, it is important to identify why retirement occurs; seven factors play a role in our models. They are (1) decreasing productivity; (2) increasing disutility of labor; (3) outside employment opportunities; (4) entitlements for retiring, such as pensions; (5) information about health and longevity; (6) indivisibility of labor; and (7) declining marginal utility of consumption. The first five factors are obviously related to the passage of time. Were they constant over time, and were labor perfectly divisible, there would be no reason to retire. Individuals would work the same amount each period. Full divisibility of labor would lead to scaled-down participation in the labor force rather than retirement. Obviously, in many instances labor is divisible at a price. One can work part time, but at a less than proportional salary. The analysis is simplified by assuming complete indivisibility of labor.

Given indivisibility of labor, individuals might still choose to work throughout their lives unless marginal utility from consumption declines. Thus, as is standard, individuals are assumed to have a concave period utility of consumption (though in some instances, to facilitate exposition, marginal utility may be constant). With decreasing productivity or increasing disutility of labor as one ages, it will be reasonable for individuals to consider retirement toward the end of their lives. If outside employment opportunities decline over time, the date of retirement from a given company will be advanced. Retirement patterns from the public employment sector illustrate this point; many individuals leave military or civil service at a time when they can still get a good outside job offer.

Entitlements may also be a function of age. Both social security and private pensions are age related. Presumably, individuals make some rough calculation of the value to them of the entitlements streams for different retirement dates and choose accordingly. A matter of central concern in this analysis is the structure of returns that an individual can expect from retirement, the effect that this will have on individual decisions, and ultimately the structure of the pension plans that will be supportable in a competitive marketplace.

In valuing the entitlements to be received on retirement, a key consideration is how long the individual expects to live. If working one more year yields a 10% increase in the per period benefit, the additional year may be a worthwhile sacrifice for a person with a life expectancy of 25 years, but not for one with an eight-year life expectancy.

Though two factors—indivisibility of labor and declining marginal utility of consumption—are in themselves sufficient to explain why individuals would take periods of leisure in their lives, they do not explain why these periods should be at the end of one's life. This phenomenon is better explained by age's adverse effects on productivity and disutility of work, in conjunction with the important role of uncertainty. The disutility of labor at later ages is something workers cannot predict accurately at an early age. They also learn more about their longevity as they age. This suggests that if the discount rates for the disutility of work and consumption are the same, it is best to work at the beginning of life and to make decisions on when to retire later when more information is available.

10.1.3 Historical Origins of Private Pensions

In 1875, the American Express Company established the first formal private pension. Only permanently disabled workers who were over 60 and had worked at least 25 years in the firm were eligible. As a large holder of railroad companies, American Express employed workers in dangerous jobs. Many of them, once injured, had no means of supporting themselves; guaranteeing workers an income if disabled made them more likely to accept dangerous jobs. By 1905, the railroads had created 12 formal plans covering 488,000 workers, or 35% of all railroad employees. By 1929, over 80% of all railroad workers were covered by some sort of retirement plan (Greenough and King 1976).

Other industries, not all in hazardous fields, also began pension plans, and by 1920 almost 400 existed. Companies began to realize the benevolent and economic consequences of retirement plans. Before long, this mostly discretionary and nonlegally binding form of retirement compension was seen as a moral obligation of the employer. Corporations welcomed this interpretation, but to become a permanent institution in the private sector, pensions had to produce some tangible economic benefits to the employer. The employee's gold watch represented an investment, not just a gift.

Most significantly, pensions enabled the employer to retire older and incapacitated workers. Previously, employers were forced to adopt such inefficient alternatives as retaining employees on the payroll at reduced pay, reassigning them to less demanding jobs, or offering occasional relief packages to particularly needy retirees. By establishing a formal pension plan, the employer could remove these workers in an orderly and employee-approved fashion without fear of adverse public reaction. When the maximum retirement age was fixed, retirement became ingrained as part of working life. Pensions allowed the employer to retire less productive workers while retaining an air of equity and appreciation for a job well done. Replacement of these older workers with younger, more agile ones would increase labor productivity through the increased efficiency of a younger work force. Pension plans thus became an instrument for fostering the retirement of older workers with a minimum of employee resistance.

While the employers emphasized the reward aspect of a pension for those actually retiring, the emphasis for the active workers was on earning the reward; it encouraged them to give the kind of service that was of greatest value to the company. By basing pensions on continuity of employment, business organizations thought they had found a means not only of preventing strikes but also of promoting long, loyal, and uninterrupted service. Presumably long and continuous service records would mean reduced labor turnover and lower training costs. Thus, pensions not only provided a means of replacing older workers with more efficient younger ones, they made younger employees more reluctant to quit and increased their efficiency; from the employers' standpoint, the economic cost of a pension could be more than offset by the improvement of labor productivity.

Pensions strengthened the worker's allegiance to the firm at the expense of his loyalty to his union. Employers understood the unions' dislike of these plans and became determined to maintain control over pensions. Over 95% of the workers covered by pension plans paid nothing into the scheme. By not requiring worker contributions and making the benefits discretionary, management could bar unions from involvement in retirement policy and set the requirements themselves. As a result, employers gained considerable leverage over employees' work decisions. The effect, it was hoped, would be a reduction in strikes and a weakening of union appeal.[1]

10.1.4 Pensions and Labor Market Performance

Many economic factors contributed to the historical development of pensions. Feldstein (1982, p. 1) identifies the myopia argument for forcing workers to protect themselves for old age when their productivity will be lower. Moreover, since pensions can be a means of backloading compensation, they may be an important factor in rewarding, motivating, and tying workers to firms. By delaying rewards, the firm can better deal with uncertainty that is resolved over time. As information on worker's effort level or productivity accumulates, the firm discovers how much to pay him. Many of the standard adverse selection and moral hazard problems disappear as the time span becomes sufficiently long (assuming that there is no efficiency cost to withholding earlier rewards).

Pensions have considerable advantages as a delayed reward mechanism. In contrast to wages that rise faster than productivity, pensions reallocate resources to a time period to which the individual himself wishes to reallocate resources. There is, of course, a limit on the amount that can be efficiently reallocated to the retirement period. Until this constraint becomes binding, pensions need not entail any efficiency loss. In this respect, they differ from other common means of withholding or backloading rewards, such as wage streams that rise faster than productivity or big prizes in contests (promotion lotteries). The function of pensions as a reward and motivational mechanism is widely cited in relation to retention of workers. But pensions may also come to play a significant role in sorting workers by quality and motivating them.

Many market imperfections may be mitigated through the use of pensions. For example, the firm may be able to invest at a greater rate of return than the worker can, whether because of tax wedges or transactions costs in raising funds. If some funds are left for the firm to invest through pensions, both parties gain. In general, this analysis looks at the benefits that go to the worker and firm together. The predominant prediction is that possible efficiencies will be pursued.

Pensions have a straightforward tax benefit as a form of compensation. They are not taxed at the time they are earned, and returns to these investments are not taxed along the way. In any single instance, it is a complex problem to figure out the precise trade-off between a dollar of pension and a dollar of compensation. But virtually all analyses agree that under current tax provisions it is desirable to use some element of pensions as part of the wage package (see Woodbury 1983).

Why should the government be promoting the use of pensions in this way? In the 1980s, some might say that the problem for our economy is insufficient capital formation. Government favoritism for pensions stretches back to periods when it was thought that insufficient consumption in the economy was the predominant problem. One possible explanation is that pensions tend to protect the government, much as flood insurance and health insurance—both subsidized—protect it. If people reach old age without a visible means of support, the government will be forced to support them.[2]

10.2 Pensions and Retirement

We have argued that pensions serve several functions distinct from their effects on retirement. This suggests that pension plans will remain part of our economy, inevitably affecting retirement, conceivably in an adverse fashion. As we observe pension plans in operation, it may be difficult to determine their intended consequences for retirement, since they serve multiple purposes. Moreover, we may not be able to tell whether their design is optimal. At some junctures, we will discuss the form of optimal pension plans given retirement objectives, stripping away other concerns. We shall also make predictions about the consequences of pensions that have traditional structures on workers' retirement decisions.

Why should pensions be used to affect retirement? A variety of reasons are identified in the models below. They center on problems stemming from an inability to make enforceable contracts. Firms typically do not reduce a worker's wages as he grows older, even though his productivity declines. If productivity could be predicted as a function of age, then in the first-best contract, the worker would agree in advance to retire at a particular age and the firm would offer a level wage over his lifetime. Even in this simplified world, where critical uncertainties about the evolution of productivity have been eliminated, present regulatory structures would make such contracting impossible. Congress recently raised the mandatory retirement age to 70. In March 1983, such "protection" was accorded as well to state and local workers. There is some speculation that prohibitions on mandatory retirement will be further relaxed.

Many firms have discovered that the best way to guard against having to keep low-productivity workers on the payroll is to offer them pension inducements to depart. Such inducements are also in the interest of workers, who would otherwise not be in a position to "promise" to retire.

Pensions play a second major role vis-a-vis retirement by spreading risks in heterogeneous populations. Suppose productivity and earnings were constant, but individuals come to differ over time in their disutility for work. The critical policy issue is how to induce those with low disutility to continue working while providing adequately for those who, say for reasons of ill health, are unable to continue productive endeavors or could do so only at unacceptable cost.

The analysis here focuses on these two concerns: inducing appropriate retirement and spreading risks associated with factors affecting retirement. In past discussion of these issues, much has been made of the question of the actuarial fairness of pension plans. Such fairness is likely to turn out to be a legal issue as well as an economic one. We believe that actuarial fairness is an elusive concept, and we discuss several different potential definitions. No single concept of actuarial fairness does a good job of capturing what we should expect from pension plans across a variety of situations.

10.2.1 The Retirement Decision

As a worker ages, several factors become important in influencing his decision to retire. Savings and pension entitlements grow with work tenure; this raises the level of sustainable consumption during retirement. This increase in consumption is further enhanced by the fact that additional work shortens the retirement period. The disutility from work also rises as the worker ages and his health deteriorates. Eventually, productivity and outside opportunities decline, although real wages may not fall to reflect this fact. Near retirement, workers have a more accurate idea of their life expectancy. They are better able to compare the trade-off between working and retirement. Other issues, such as the presumed improvement in the quality of life and increased longevity from retirement, are also important when making the retirement decision.

The incentives to retire increase with age. The main factors influencing retirement are shown in figure 10.1. Here $\tilde{L}(R)$ = disutility of work at age R; $W(R)$ = wage at R; $C_2(R)$ = retirement consumption given retirement at age R; $F(R)$ = productivity at age R.

The individual should continue working until the utility of working plus the higher retirement consumption it affords falls short of the utility of retirement.

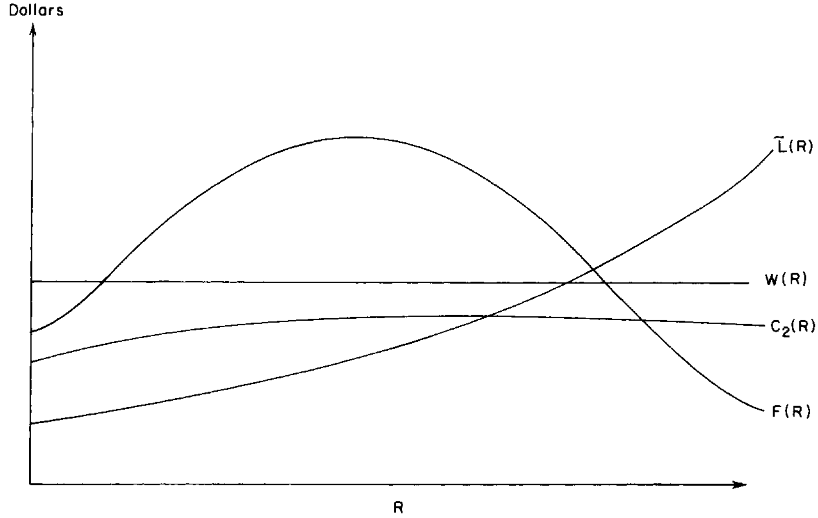

Fig. 10.1

10.2.2 Savings and Appropriate Retirement

The decision to retire is based largely on a worker's postretirement income and his perceived health and life expectancy. Workers with sufficient foresight can save for themselves. Alternatively, their employer can provide forced savings as part of the total compensation package. In the second-best world of labor contracting, it is not always advantageous for workers to have the option of choosing their level of savings. When a worker cannot commit himself to retire at the optimal age, pensions can induce him to retire earlier. By forcing a worker to save too much, pensions offset the externality created by the fact that wages do not fall during an older employee's less productive years.

Firms base wages on the expected productivity of a worker over his lifetime with the firm. The wage schedule is decided in advance and is based on a worker's expected retirement age. Given the schedule, the worker is then free to choose when to retire. But, his actions must be consistent with what the firm expected him to do.

Because wages stay constant over a worker's lifetime, he will not have sufficient incentives to retire when productivity is declining. Laws against mandatory retirement prohibit contractual agreements to retire at a prespecified age, and firms realize that, without a binding commitment, workers will delay their retirement. Accordingly, they reduce lifetime wages to compensate for the period of high wages and low productivity before retirement.

Pensions help solve this problem by providing economic incentives to retire at an earlier age.[3] Pensions put aside higher savings than the worker would choose and make them accessible only on retirement. They provide the counterbalance that speeds up retirement that otherwise would be suboptimally delayed by wages higher than productivity.

Under any given wage schedule, an employee would be better off if he could select both his savings and his retirement age. But the same wage schedules are then not feasible. Workers who feel that they are being forced to save too much are not attracted by contracts without pensions because the freedom to save is more than offset by the resulting lower wages (due to employer's inability to induce retirement).

We illustrate this use of pensions in a simple model in which workers care only about consumption and leisure. They start with a firm at age Y_0 and stay with the firm until they retire at age R. Wages are constant and equal to expected productivity. A worker of age Y has productivity $F(Y)$. The risk-neutral employer, if he is to attract any workers in a competitive equilibrium, must maximize his workers' expected utilities subject to the zero profit constraint. Initially, we assume that the discount rate is zero for both workers and firms,[4] and that everyone knows his exact life span. The results can be extended to include positive discounting and uncertain life expectancy under the assumption that perfect annuity markets exist.

The worker's period utility is represented by $U(C, L)$,

$$(1) \qquad \partial U(C,L)/\partial C > 0; \quad \partial^2 U(C,L)/\partial C^2 < 0; \quad \partial U(C,L)/\partial L < 0,$$

where C is consumption and L is labor supply. Institutional requirements force the labor supply to be either $\bar{L}$ or 0. The age-productivity profile is assumed first to rise as the worker gains experience and eventually to fall if the worker stays past a sufficiently old age.

It is optimal for consumption levels to be constant, given the labor supply, as there is declining marginal utility from consumption. The worker chooses his consumption while employed, C_1, and his consumption when retired, C_2. Since both wages and consumption (while working) are constant, savings will also be a constant fraction, α, of salary. The retirement decision, R, is constrained by the condition that savings must be sufficient to meet the cost of the expected retirement consumption. Total lifetime utility is given by

$$(2) \qquad (R - Y_0)U(C_1, \bar{L}) + (T - R)U(C_2, 0)$$

Consumption is financed out of earnings and savings. The budget constraint for a worker earning wage W can be separated into two parts, work life and retirement. When working, his consumption is by definition equal to his salary net of savings,

$$(3) \qquad C_1 = W(1 - \alpha).$$

The retirement age and postretirement consumption are jointly determined. Given a desired level of postretirement consumption, a worker retires when his accumulated savings are sufficient to finance his consumption. For a worker who knows that his life span is exactly T years, this implies:

$$W\alpha(R - Y_0) = (T - R)C_2. \tag{4}$$

A worker's lifetime wage is based on his expected productivity,

$$W = \left[\int_{Y_0}^{R} F(Y)dY \right] / (R - Y_0). \tag{5}$$

Assuming competitive conditions, a firm knowing a worker's R would offer him the wage defined by (5). If the worker could commit himself to retire at a prespecified time, he would select R to maximize his utility, taking into account the effect on his wages of postponing retirement. (If there were uncertainties to unfold in the future, he would make contingent retirement commitments contingent on his condition.)

Unfortunately, a worker cannot commit himself to retire at a prespecified age. His wages must be determined in advance, and independently from his actual retirement decision. Moreover, if productivity declines later in life, as is commonly the case, wages are likely to exceed productivity. A worker deciding on retirement will equate marginal disutility with wage, whereas efficiency requires that it be equated with productivity. This constellation of factors creates an inefficient situation. Workers, unable to commit themselves in advance to a retirement date, will choose to retire "too late."

The firm meets its zero profit constraint by choosing wages based on when it expects its employees to retire. These assumptions must be consistent with the optimal retirement age given the chosen wages. Given a fixed wage W, a worker chooses C_1 and C_2 to maximize his lifetime utility subject to (3) and (4). The Lagrangian may be written as

$$\max_{C_1, C_2, \alpha, R} \mathcal{L} = U(C_1, \bar{L})(R - Y_0) + U(C_2, 0)(T - R) + \lambda[W(1 - \alpha) - C_1] + \mu[W\alpha(R - Y_0) - C_2(T - R)]. \tag{6}$$

The constrained optimal choice of C_1, C_2, α, and R yields

$$C_1: \; U_1(C_1, \bar{L})(R - Y_0) - \lambda = 0 \tag{7}$$

$$C_2: \; U_2(C_2, 0)(T - R) - \mu(T - R) = 0 \tag{8}$$

$$\alpha: \; -\lambda W + \mu W(R - Y_0) = 0 \tag{9}$$

$$R: \; [U(C_1, \bar{L}) - U(C_2, 0)] + \mu(W\alpha + C_2) = 0, \tag{10}$$

where $U_1(C, L) = \partial U(C, L)/\partial C$.

The equations may be combined to provide two first-order conditions that are more intuitive:

$$U_1(C_1, \overline{L}) = U_1(C_2, 0), \tag{11}$$

$$U_1(C_2, 0) = \frac{U(C_2, 0) - U(C_1, \overline{L})}{W\alpha + C_2} \tag{12}$$

Because there are perfect capital markets, resources will be transferred from the work period to the retirement period until the marginal utilities of consumption are equalized. Working longer brings an extra $W\alpha + C_2$ in savings. This must be balanced by the loss in utility from delaying retirement.

The optimal solution is determined by jointly solving equations (3), (4), (10), (11), and (5); this yields C_1^*, C_2^*, α^*, R^*, and the resulting W. If, at the retirement age, wages are more than productivity, $F(R^*) > W$, then workers will have incentives to work for an inefficiently long time. The firm can partially correct this problem by forcing workers to save more.[5] In theorem 1, forced saving is shown to induce earlier retirement, and this results in higher wages.

Since C_1, C_2, α, and R are all chosen optimally, small changes in their values will not directly affect expected utility. The only way to improve expected utility is to be able to support higher wages,

$$\begin{aligned} \frac{dEU}{dW} &= \frac{\partial EU}{\partial C_1}\frac{dC_1}{dW} + \frac{\partial EU}{\partial C_2}\frac{dC_2}{dW} + \\ &\quad \frac{\partial EU}{\partial \alpha}\frac{d\alpha}{dW} + \frac{\partial EU}{\partial R}\frac{dR}{dW} + \frac{\partial EU}{\partial W} \\ &= \lambda(1 - \alpha) + \mu\alpha(R - Y_0) \\ &= U_1(C_1, \overline{L})(R - Y_0) > 0. \end{aligned} \tag{13}$$

In equilibrium, as determined by equation (5), the wage rate is a function of the retirement age and productivity,

$$\frac{dW}{dR} = \frac{F(R) - W}{R - Y_0}. \tag{14}$$

Theorem 1: Under the assumption that $F(R) < W$, increasing the savings rate, α, above α^* improves expected utility, $(dEU/d\alpha|\alpha^*) > 0$.

Proof: From the assumptions of the theorem, earlier retirement raises wages, $(dW/dR) < 0$. Higher wages improve expected utility (eq. [13]). Thus, it only needs to be demonstrated that raising α hastens retirement. This part of the proof is longer and more complicated. Differentiating the budget constraints (3) and (4) shows

$$\frac{dC_1}{d\alpha} - (1 - \alpha)\frac{dW}{d\alpha} + W = 0, \tag{15}$$

(16)

$$\frac{dC_2}{d\alpha}(T - R) - (W\alpha + C_2)\frac{dR}{d\alpha} - \alpha(R - Y_0)\frac{dW}{d\alpha} - W(R - Y_0) = 0.$$

As W is affected by α only through R, we use equation (14) to substitute

$$-\alpha[R - Y_0]\frac{dW}{d\alpha} = \alpha[W - F(R)]\frac{dR}{d\alpha} \tag{17}$$

into equation (16). Collecting terms, rewrite equation (16) as

$$\frac{dR}{d\alpha}[\alpha F(R) + C_2] = \frac{dC_2}{d\alpha}(T - R) - W(R - Y_0). \tag{16'}$$

To determine $dC_2/d\alpha$, differentiate equation (12),

$$\frac{dC_2}{d\alpha}[U_{11}(C_2, 0)] = -\frac{[U_1(C_1, \bar{L})\frac{dC_1}{d\alpha} + U_1(C_2, 0)(W + \alpha\frac{dW}{d\alpha})]}{[W\alpha + C_2]} \tag{18}$$

Note from equation (11) that $U_1(C_1, \bar{L}) = U_1(C_2, 0)$. Thus, substitute the value of $dC_1/d\alpha$ from (15) and collect terms,

$$\frac{dC_2}{d\alpha} = \frac{\frac{dW}{d\alpha}}{(W\alpha + C_2)A} = \frac{[F(R) - W]\frac{dR}{d\alpha}}{(R - Y_0)(W\alpha + C_2)A}, \tag{19}$$

where $A = -U_{11}(C_2,0)/\ U_1(C_2, 0)$ is the measure of absolute risk aversion. It is now possible to sign $dR/d\alpha$ using (16′), (19) and (4),

$$\frac{dR}{d\alpha} = \frac{-W(R - Y_0)}{\alpha F(R) + C_2 + W\alpha[W - F(R)]/[C_2(W\alpha + C_2)A]} < 0. \tag{20}$$

Changes in α affect expected utility only through changes in W. The argument above demonstrates that when savings are "too high," retirement takes place earlier, wages are higher, and workers are better off. Essentially, when choosing α^* and R^*, workers neglect the impact on their wages. This externality is reduced if the firm chooses $\alpha > \alpha^*$; the worker is induced to retire earlier and expected utility increases.

If firms are providing pensions that are larger than workers' desired savings, then why do we observe any private savings taking place? Savings in the form of pensions is not a perfect substitute for other types of savings. In particular, since pensions are accessible only on retirement after the age of 55, they cannot provide capital needed for large purchases (such as a house) or insurance against preretirement events such as illness or unemployment. Some forms of savings outside of pensions—notably investment in home ownership—are encouraged and subsidized by the government. The value of their house, net of mortgage, forms the largest part of most families' savings.

10.2.3 Life Expectancy and Retirement—Who Gains, Who Loses

A major role of pensions is to induce workers to retire. This role will become even more important if the current movement against mandatory retirement succeeds. Pensions may provide economic incentives to retire in two significant ways: (1) pensions provide savings that can be accessed only during retirement; (2) pensions may not be actuarially fair to workers who retire after normal retirement age. The forced savings role of pensions is discussed in the previous section. This section concentrates on the actuarial value of the pension as a function of the retirement date.

Workers in their later years earn wages greater than their productivity and thus have incentives to work after their first-best retirement date. Lazear (in this vol.) argues that it is possible to correct for this externality by reducing the actuarial value of pensions to workers who postpone retirement. Changing the compensation through pensions is a graceful way to lower the "effective" wage (salary plus pension value) and restore incentives to retire early.

A first look at the data seems to confirm this observation (see the appendix). Most defined benefit plans reduce benefits only by 4%–6% for each year of early retirement (before age 65). For workers retiring after 65, pensions are generally not increased (except to take account of extra years of service). If all workers had identically distributed life spans, then, as illustrated in the model below, an actuarially fair benefit reduction for early retirement would be close to 9.5% and benefits would be similarly increased for late retirement.[6] This significantly larger factor reflects the fact that early retirement gives an extra year of benefits now; the costs occur over a discounted and uncertain future and are proportionally less important.

Consider a worker with pension P per year that is reduced (increased) by fraction b for each year of early (late) retirement. Workers know only the probability distribution of their death date. The real discount rate is 3%.[7] No retirement is allowed before age 55. At age 55, a worker's chance of living until age $55 + r$ is $G(r)$. With a uniform distribution of life spans between 55 and 90, $G(r) = 1 - r/35$. The equality defining an actuarially fair adjustment for retirement at age $55 + r - 1$ is

$$(21)\quad [1 - b(r)]P \int_{r-1}^{35} e^{-.03t}(1 - t/35)dt = P \int_{r}^{35} e^{-.03t}(1 - t/35)dt$$

This implies

$$(22)\quad b(r) \int_{r-1}^{35} e^{-.03t}(35 - t)dt = \int_{r-1}^{35} e^{-.03t}(35 - t)dt.$$

At age 65, $r = 10$ and the appropriate penalty for early retirement is

$$(23)\quad b(r) = [(33.3t - 55.5)e^{-.03t}|_{9}^{10}]/[(33.3t - 55.5)e^{-.03t}|_{9}^{35}] \approx .095.$$

Results above suggest that with most penalties smaller than 6% per year of early retirement (and no bonus for late retirement), workers who retire late receive relatively less in pension benefits than those who retire early.

Appearances may be deceiving. When workers have different life spans and know these differences, then the decision about when to retire is highly correlated with age. Workers who know that they have relatively longer life spans work later into their life. The literature of gerontology suggests that workers who retire early have shorter life spans. For example, the work of Haynes et al. (1979) demonstrates that the mortality rate of early retirees is higher than would be expected if no self-selection were occurring.

The fact that pensions do not seem to rise very fast with retirement age may reflect the different life expectancies for the different retiring age groups. Indeed, given the structure and parameters of most defined benefit plans, workers who postpone retirement receive relatively more benefits than those who retire earlier. Otherwise there would be a severe problem of adverse selection. If workers who retire early were given larger benefits to compensate them for their shorter life expectancy, then workers with long life expectancies would also retire early.[8] The pension plan could not afford to pay pensions based on short life expectancies to workers with long life spans who retire early.[9] The model presented below shows that workers who retire early are actuarially penalized in a typical defined benefit pension plan. This may be necessary to give workers with longer life expectancies sufficient incentives to remain in the labor force.

The typical defined benefit pension plan has payments proportional to a function of wages[10] multiplied by years of service with a multiplicative linear penalty for early retirement. A worker entering at age Y_0 and retiring at R receives an annual pension

$$(24)\qquad P(R) = \begin{cases} \tilde{\alpha}F(W)[R - Y_0][1 + (R - \bar{R})\beta], & Y \le \bar{R}, \\ \tilde{\alpha}F(W)[R - Y_0], & Y > \bar{R}. \end{cases}$$

When a firm can choose the fraction of wages, $\tilde{\alpha}$, normal retirement date, $\bar{R}$, and the early retirement penalty, β, then the pension payment for retirement before age $\bar{R}$ can be written as a general quadratic function of R,

$$(25)\qquad P(R) = A(R^2 + BR + C)$$

where $A = -\tilde{\alpha}F(W)\beta$; $B = [1 - \beta(\bar{R} + Y_0)]/\beta$; $C = Y_0(\bar{R}\beta - 1)/\beta$.

The pension payment for retirement after age R is a linear function of $\bar{R}$ as seen in equation (24). There is some loss in generality when wages are not also a function of retirement age and seniority. However, at least for pattern plans, $F(W)$ is a constant and this effect is unimportant. At present, we are concerned with demonstrating the effect of life span on retirement decisions and the actuarial value of pensions, and hence assume that wages and productivity are constant over the life span. Wages are not affected by the retirement date. The discount rate is zero and conditional on the labor supply, the marginal utility of consumption is constant. Thus, retirement decisions will be based solely on expected longevity.

A worker who knows that his life span will be exactly T years and who faces a pension schedule, $P(R)$, chooses his retirement age R to maximize his expected lifetime utility.

$$EU = U[(1-\alpha)W, \bar{L}](R - Y_0) + U[P(R), 0](T - R). \tag{26}$$

The first-order condition determining the optimal retirement date is

$$P'(R)(T - R)U_1[P(R), 0] = U[P(R), 0] - U[(1-\alpha)W, \bar{L}]. \tag{27}$$

For retirement prior to age $\bar{R}$ this equation can be solved explicitly for R as a function of T under the earlier assumption of a quadratic pension plan and constant marginal utility of consumption,

$$(2AR + AB)(T - R)U_1(0,0) = A(R^2 + BR + C)U_1(0,0) - (1-\alpha)WU_1(0,\bar{L}) + U(0,0) - U(0,\bar{L}). \tag{28}$$

Let γ equal the ratio of the marginal utility of consumption when retired to when employed, $\gamma = U_1(0,0)/U_1(0,\bar{L})$. Define $\tilde{L}$ as the disutility from work measured in wage units, $[U(0,0) - U(0,\bar{L})]/U_1(0,0)$. Then, (28) can be reduced to

$$R(T) = \frac{1}{3}\left(T - B + \sqrt{T^2 + B^2 + TB - 3\{C + [\tilde{L} - (1-\alpha)W\gamma]/A\}}\right) \tag{29}$$

For retirement after age $\bar{R}$, pension benefits rise linearly with age. The optimal retirement date as a function of life span is

$$R(T) = \frac{1}{2}\{T + Y_0 + [(1-\alpha)\gamma - \tilde{L}]/[\tilde{\alpha}F(W)]\}. \tag{30}$$

Usually, there will be some overlap among these solutions; during this period all workers choose to retire at age $\bar{R}$. A small difference in the marginal incentives for retirement before versus after $\bar{R}$ may induce a large segment of the work force to retire at $\bar{R}$.

To illustrate an example of the optimal retirement decisions, let pension benefits be 1.5% of net wages for each year of service with a 5% reduction for each year of early retirement before age 65: $\tilde{\alpha} = .015$, $\beta = .05$, $F(W) = (1-\alpha)W$, and $\bar{R} = 65$. The disutility from work is $.57W$. Saving 14% of wages balances the budget when population life spans are triangularly distributed between ages 59 and 88. The solution to the optimal retirement decision is graphed below in figure 10.2.

Who wins and who loses? The value of a pension to a worker with life span T is

$$V(T) = P[R(T)][T - R(T)] - \alpha W[R(T) - Y_0]. \tag{31}$$

Living longer changes this value by

$$V'(T) = P[RC(T)] - R'(T)(W - \tilde{L}), \tag{32}$$

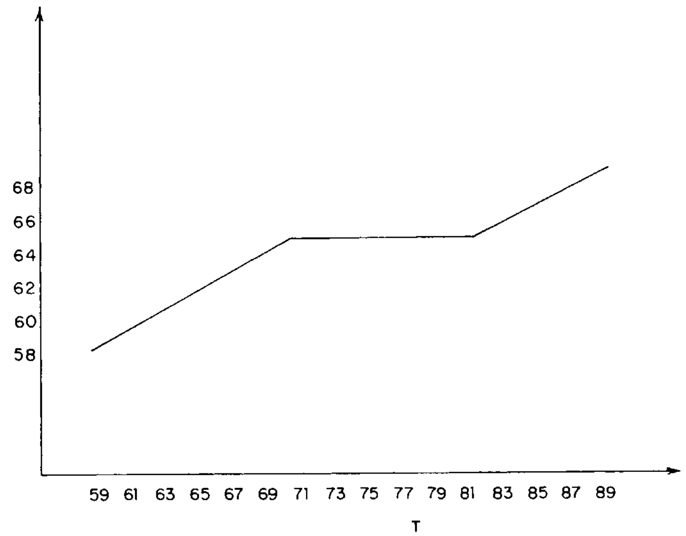

Fig. 10.2 Retirement age as a function of life span.

where $P'(R)$ was substituted in from equation (27). Workers who live longer have greater pension benefits provided that their replacement ratio, P/W, exceeds their incremental time in the labor force, R', times the net dollar value of working, $1 - \tilde{L}/W$. In the example presented above, pension value strictly increases with life span (and hence retirement age) as illustrated in figure 10.3.

10.3 Optimal Pensions with Heterogeneous Populations

What are the consequences of offering a single pension plan to a heterogeneous population? Models in various areas, ranging from insurance markets to health coverage to labor contracting, have shown that important problems arise when a single policy is applied to a heterogeneous population. Pensions in some sense combine elements of all of these models.

Heterogeneity would pose no difficulty if it could be readily diagnosed and if different policies could be offered to workers with different characteristics. Such screening is impossible in practice for a variety of reasons. The differences among workers may be imperceptible to the employer. Moreover, the pension contract is generally drawn up long before many of the differentiating characteristics manifest themselves to anyone, including the worker. It would be a violation of the earlier contract to exclude any worker from a pension option.[11] Legal restrictions and general labor practices make it exceedingly difficult to afford different treatment to workers who, despite differences in some present characteristics, have the same employment histories. Unless pensions are negotiated through col-

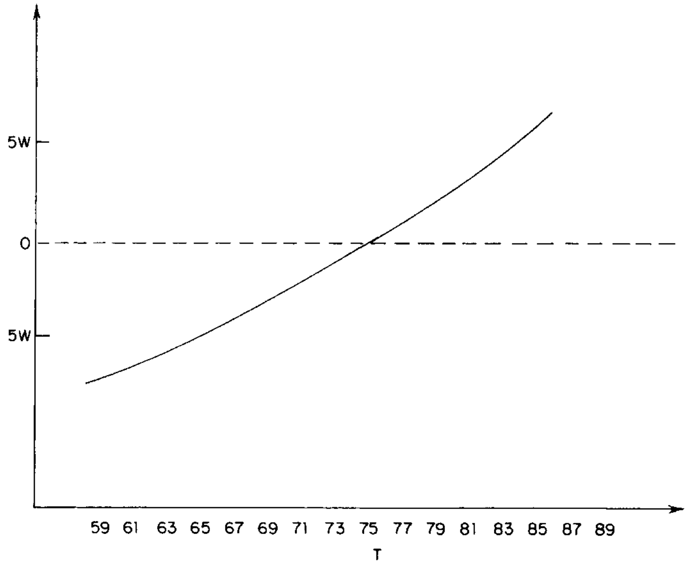

Fig. 10.3 Net actuarial value as a function of death date.

lective bargaining, corporations are prohibited from offering different pensions to different classes of workers. Thus, a corporation is not permitted to have different pension plans for white-collar and blue-collar workers.

10.3.1 The Order of Retirement

The decision to retire is determined by many factors; the following models focus on the role of life expectancy and disutility from work. All other things equal, workers who live longer or who enjoy work more will choose to retire later. This result is true for any pension plan whose annual payments do not decrease with retirement age.

Consider workers of type $i = (A,B)$ who enter the labor force at age Y_i and retire at age R_i with pension $P(R_i)$ and life expectancy T_i. Utility when employed is $U(C_i, i)$ and when retired is $U[P(R_i),0]$. Assume that it is optimal for A to retire before $B, R_A < R_B$. Because A retires earlier, his annual pension payment is smaller, $P(R_A) < P(R_B)$.

Theorem 2: Under the assumptions stated above, worker B, who retires later, must have either a longer life expectancy, a greater utility when employed, or both.

Proof: The fact that each worker prefers his retirement date implies

$$(33) \quad \begin{aligned} &U[P(R_A), 0](T_A - R_A) + U(C_A, A)(R_A - Y_A) \\ &\geq U[P(R_B), 0](T_A - R_B) + U(C_A, A)(R_B - Y_A), \end{aligned}$$

$$(34) \quad \begin{aligned} &U[P(R_B), 0](T_B - R_B) + U(C_B, B)(R_B - Y_B) \\ &\geq U[P(R_B), 0](T_B - R_A) + U(C_B, B)(R_A - Y_B). \end{aligned}$$

Adding these inequalities together yields

$$(35) \quad \begin{aligned} &(R_A - R_B)[U(C_B, B) - U(C_A, A)] \\ &\geq \{U[P(R_B), 0] - U[P(R_A), 0]\}(T_A - T_B). \end{aligned}$$

Any contradiction to the theorem would require the worker who retires earlier to have both a longer expected life span and greater utility from work, $T_A > T_B$ and $U(C_A, A) > U(C_B, B)$. However, this must violate equation (35).

10.3.2 The Costs of Postponing Retirement under Optimal Pensions

If two groups differ in terms of either length of life or utility from work, one group will always retire before the other. This no-switching property results more generally when the indifference curves of different groups have a single crossing point (SCP).[12] This seems to be an especially reasonable assumption in the context of retirement decisions. We further assume that all workers have the same utility function when retired. There is neither reemployment nor outside income once retired. Moreover, given that retirement is preferable to working, the schedule of annual pension payments must be increasing with age, otherwise workers would retire earlier. The higher annual pension payments to workers who retire later imply that they have a lower marginal utility of consumption.

Here we shall seek the optimal pension scheme; it is the benefit schedule that maximizes the sum of the individuals' utilities subject to a budget constraint. A salient feature of the optimal pension schedule is that it is actuarially unfavorable for a specified individual to retire later. The logic supporting this result parallels that in the optimal taxation literature: it is desirable to tax those with lower marginal utilities of income (penalize late retirees in this instance) to redistribute income to those with higher marginal utility of income (early retirees). At least initially, efficiency loss associated with inappropriately influencing work or retirement are outweighed by the utility gains from the income transfer. This result is formalized in theorem 3.

Theorem 3: The SCP property is sufficient to imply that with the optimal pension scheme, the value of pensions plus wages net of productivity must be actuarially larger for a worker who chooses to retire earlier; that is, any particular worker who retires earlier will receive pensions plus total wages (net of productivity) with a higher actuarial value.

Comment: It is still possible that the type of workers who choose to retire earlier may be the ones with lower pensions; indeed, as seen in section 10.2.3, current pension plans actuarially favor later retirees. Thus, in the subsequent discussion it is important to distinguish between the differences in the actuarial value of a particular worker's pension based on his decision when to retire (individual fairness) and the differences between the actuarial value of a pension for different types of workers (cohort fairness). It is also important to maintain the assumption that reemployment once retired is not permitted; otherwise, workers would retire if their pension value began to decline and then seek reemployment.

Proof: In the proof, we take productivity and wages as given; pensions are adjusted to achieve the result. The firm offers the worker a choice between various retirement dates and their associated pension benefit, $[R,P(R)]$. The first part of the proof demonstrates that each worker must be indifferent between his retirement date and the one preceding his. Given indifference, earlier retirement is shown to cost the firm more. Intuitively the optimal pension scheme will distribute the largest feasible benefits to workers retiring early since they receive the lowest payment and thus have the highest marginal utility of income. Because early retirees' pensions are maximized, late retirees will be pushed to the point of indifference between their chosen retirement date and retiring earlier. However, if the late retirement benefit had a higher actuarial value, the pension plan could improve welfare by eliminating the expensive option; the late retiree is indifferent to retiring earlier so that his utility is the same but the total cost to the firm is smaller.

Consider the first worker to retire who strictly prefers his retirement date to the one preceding his. This is the case with worker C illustrated in figure 10.4. Lower his pension and raise the benefits for worker B. Since C strictly preferred retiring at R_C rather than R_B, it is possible to raise $P(R_B)$ and lower $P(R_C)$ without inducing C to retire earlier. Raising $P(R_B)$ does not result in A's retiring later because A must strictly prefer retiring at R_A to R_B. This strict preference follows from the single crossing property and the fact that B is indifferent between retiring at R_A and R_B (as C was the first worker to strictly prefer his retirement date over the preceding one). By transferring money from worker C to B, social welfare is improved; worker B has a smaller pension and thus a higher marginal utility of income. An optimal pension scheme must make workers indifferent between their retirement date and the preceding retirement date.

Since workers are indifferent between their chosen retirement date and the preceding one, it directly follows that the later (and chosen) retirement date must have a lower total cost to the firm. If not, simply eliminate the later retirement date. The worker then takes the earlier retirement; he is indifferent. His utility remains constant, but the total cost to the pension fund is smaller. The excess profits can be redistributed to make everyone better off.

This result holds for a broad range of problems. Workers may have varying life spans, differing disutilities from work, and unequal productivities. However, in all these problems the marginal utility of income is a

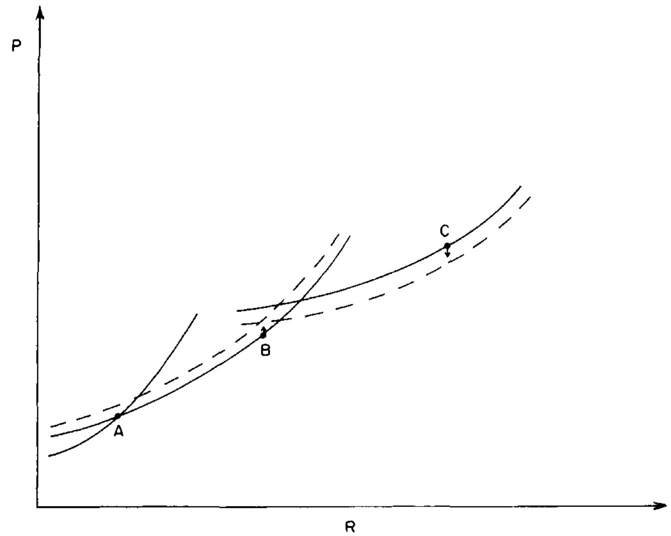

Fig. 10.4

decreasing function of retirement age; workers who retire later receive larger pension payments and thus have smaller marginal utilities of income. The theorem would be reversed if the marginal utility of income rose as a function of retirement age; then, the value of pension payments plus wages net of productivity would be more than actuarially fair for workers who delayed their retirement. This situation might occur if the primary difference among individuals is outside income. If those who retire early are the ones with high outside income and thus low marginal utility of income, then the optimal pension scheme (plus wages net of productivity) must be actuarially advantageous to a worker who chooses to retire later. Theorem 3 has a simple generalization.

Theorem 3′: The SCP property is sufficient to imply that with the optimal pension scheme, the value of pensions plus wages net of productivity must be actuarially larger for a worker who chooses to retire earlier (later) if marginal utility of consumption decreases (increases) with retirement age.

The proof for theorem 3′ needs only one small modification to incorporate the generalization that the marginal utility of consumption may increase with retirement age. In this case, workers will all be indifferent between their retirement date and the following one. The earlier retirement date cannot be more expensive as otherwise it will be eliminated.

An illustration of theorem 3, that is, when marginal utility of income declines with retirement age, is presented in the following section.

10.3.3 Disutility and Optimal Pension Design

Disutility of work generally increases with age. Given institutional constraints against shortened work weeks, eventually the disutility of labor becomes large enough to induce retirement (see, e.g., Hausman and Wise in this vol.). Workers would like to be able to insure themselves against prematurely having a high disutility from work. This problem is most severe when the worker is actually disabled and can no longer work. A disability is an extreme form of disutility from work. To illustrate an application of theorem 3, this section presents a simplified version of a model initially studied by Diamond and Mirrlees (1978). Pensions are used to support disabled workers, an extreme case that simplifies exposition. Qualitatively these results apply to the range of situations where the decision to retire is based on the level of disutility from work.

Pensions, as they were originally conceived, can be viewed as insurance to workers who are disabled and forced to retire (or have high disutility and would choose to do so). However, a problem arises when the disability is not observable or verifiable. Healthy workers must have sufficient incentives to continue working. Consequently, less insurance is provided than in the first-best solution when the disability is observable. Insurance can still be provided to workers who can prove that they are disabled—witness the separate provisions for early retirement with certain restricted and verifiable disabilities.

To highlight the insurance feature of pensions, let the disability affect neither the utility of consumption nor the worker's life expectancy.[13] There may be some exogenous utility loss but, effectively, the disability simply forces the worker to retire. It is easiest to imagine that the disutility from labor is initially zero and becomes infinite if the worker remains in the labor force after he is disabled.

In a first-best contract, pensions are conditional on the worker becoming disabled and there is no problem of adverse selection. Disabilities occur stochastically. Pensions should be constant, independent of the retirement date. This is demonstrated below.

Let utility from income Y be $U(Y)$, whether working or not. The discount rate is zero. The disability takes place at age R with probability distribution $h(R)$ and is verifiable. Workers all have a life expectancy of exactly T years. The savings and contributions to pensions have previously taken place; there is a fixed amount of savings, S, available to fund the pension scheme (and no other private savings are allowed). While working, employees are given their constant marginal product, W. The optimal pension payment, $P^*(R)$, is constant and independent of the retirement date. There is, however, a maximum retirement date, $\hat{R}$. Even workers who are healthy should retire by $\hat{R}$ in order to receive some benefit from the pension savings. A worker's utility is maximized subject to the constraint that the expected pension payout equal the pension savings,

$$\frac{Max}{\hat{R},\, P(R)} \mathcal{L} = \int_{Y_0}^{\hat{R}} \{U(W)(R - Y_0) + U[P(R)](T - R)\} h(R) dR$$

$$(36) \qquad + \{U(W)(\hat{R} - Y_0) + U[P(\hat{R})](T - \hat{R})\}[1 - H(\hat{R})]$$

$$+ \lambda \left\{ S - \int_{Y_0}^{\hat{R}} P(R)(T - R)h(R)dR - P(\hat{R})(T - \hat{R})[1 - H(\hat{R})] \right\}$$

The first-order conditions are

$$(37) \qquad U'[P(R)](T - R)h(R) - \lambda(T - R)h(R) = 0 \rightarrow U'[P(R)] = \lambda,$$

$$(38) \qquad \{U(W) - U[P(\hat{R})]\}[1 - H(\hat{R})] + \lambda P(\hat{R})[1 - H(\hat{R})] = 0.$$

Together they imply that the pension is constant and

$$(39) \qquad U[P(\hat{R})] - U(W) = U'[P(\hat{R})]P(\hat{R}).$$

Depending on the level of savings, S, there are three possible solutions to the first-order conditions. The maximum retirement date is chosen to make the constant pension, $P(\hat{R})$, affordable. If this is possible, the level of pensions does not depend on the amount of pension savings. As the savings become larger, the maximum retirement date is simply shifted forward. It is conceivable that the savings are sufficiently small that $P(\hat{R})$ is never affordable; then, there is no maximum retirement date ($\hat{R} = T$) and the pension payments are just large enough to exhaust the budget, $P = S/(T - \bar{R})$ where $\bar{R}$ is the expectation of the disability date. At the other less realistic extreme, when savings are larger than needed to finance $P(\hat{R})$, everyone retires immediately and the savings are equally distributed, $P = S/(T - Y_0)$.

It is impossible to provide the same constant pension scheme when disabilities are not verifiable. All workers would choose to retire at the earliest possible age. The problem is then to give high pensions to workers who are disabled at a young age without inducing other healthy workers to retire. The optimal second-best solution is obtained when the feasible pension for the youngest disabled worker is maximized. As seen in the first-best solution, the ideal pension is equal for all workers. When pension payments are unequal, welfare can be improved by transferring pension wealth from workers who receive high payments to workers who receive lower payments. The problem of adverse selection results in workers who are disabled young receiving the lowest pension. To the extent that it is feasible (in terms of both the budget constraint and the self-selection constraint) to raise the payments to young disabled workers, welfare is improved.

In the constrained optimal solution, a healthy worker must be indifferent between retiring and working (as proven in Diamond and Mirrlees [1978]). This holds if pension benefits satisfy the relationship,

$$(40) \qquad U(W) - U[P(R)] + U'[P(R)]P'(R)(T - R) = 0.$$

This differential equation determines the minimum increase in pensions with retirement age necessary to prevent healthy workers from pretending that they are disabled. Under this formulation, a worker's retirement age is a matter of indifference, and a worker is assumed to continue working until he becomes disabled. Because of the indifference to retirement date, utility at the time of retirement when disabled is identical to the utility at the minimum retirement date,[14] R_0, with pension $P(R_0)$,

$$\begin{aligned} U(W)(R - Y_0) + U[P(R)](T - R) \\ \equiv U(W)(R_0 - Y_0) + U[P(R_0)](T - R_0). \end{aligned} \tag{41}$$

Since the utility is identical for all workers, expected utility is also equal to this representative utility. As can be seen from equation (41), expected utility increases with $P(R_0)$.

The pension payments are chosen to rise just fast enough to keep workers from retiring until they are disabled. There is also a maximum retirement age, R^*. Because pensions are constant from R^* onward, at age R^*, all remaining workers will choose to retire, disabled or not. The budget constraint requires that the cost of the pensions, $B[P(R_0),R^*]$, must equal the pension savings, S, where

$$\begin{aligned} B[P(R_0), R^*] &= H(R_0)P(R_0)(T - R_0) \\ &+ \int_{R_0}^{R^*} P(R)(T - R)h(R)dR \\ &+ [1 - H(R^*)]P(R^*)(T - R^*). \end{aligned} \tag{42}$$

The choice of R^* leads to an R_0, which in turn determines expected utility. Recall that $P(R)$ is determined by the initial conditions and equation (40).

$$P(R) = \begin{cases} P(R_0) + \int_{R_0}^{R} \{U[P(x)] - U(W)\}/\{U'[P(x)](T - x)\}dx, & R_0 \le R \le R^* \\ P(R^*) & R > R^* \end{cases} \tag{43}$$

Increasing $P(R_0)$ raises $P(R)$ for R between R_0 and R^*. Hence, the cost of an increase in $P(R_0)$ is unambiguously positive.

$$\partial B[P(R_0), R^*]/\partial P(R_0) > 0. \tag{44}$$

It is possible to increase $P(R_0)$ and satisfy the budget constraint only so long as $\partial B[P(R_0), R^*]/\partial R^* < 0$, where

$$\frac{\partial B[P(R_0), R^*]}{\partial R^*} = [1 - H(R^*)][P'(R^*)(T - R^*) - P(R^*)] \tag{45}$$

$$= \frac{[1 - H(R^*)]\{U[P(R^*)] - U(W) - P(R^*)U'[P(R^*)]\}}{U'[P(R^*)]}. \tag{46}$$

To maximize $P(R_0)$, the optimal choice of R^* occurs at the unique solution[15] to $\partial B/\partial R^* = 0$,

$$U[P(R^*)] - U(W) - P(R^*)U'[P(R^*)] = 0. \tag{47}$$

Because pensions rise with retirement age, workers receive less than perfect insurance. Workers who are disabled later benefit at the expense of those who must retire early. On the other hand, theorem 3 still applies; pensions rise more slowly than is actuarially fair.

Right at the maximum retirement age, R^*, there is no longer any adverse selection and pensions rise exactly at the actuarially fair rate. From equation (46), $P'(R^*) = P(R^*)/(T - R^*)$. A comparison of second derivatives[16] shows that the actuarially fair curve is steeper than $P(R)$ for retirement before R^*. Second-best pensions must be as illustrated in figure 10.5.

As an example of the second-best solution, consider workers who arrive at age 60 each with 60 units of pension savings. Life span is known to be exactly 20 years. The probability of disability is uniform between ages 60 and 80. While the individual works, his wages are 2. Utility is logarithmic and there is no disutility from work until disabled. The first-best pension plan (when disabilities are verifiable) offers an annual benefit of 5.4 to disabled workers and permits everyone to retire after 13.67 years. When disabilities are unobservable, the constrained optimal pension offers an annual benefit of 3.3 for retirement at age 60 and rises exponentially to a maximal payment of 5.4 for retirement at age 70. Between ages 60 and 70, only disabled workers retire; at age 70, all remaining workers retire.

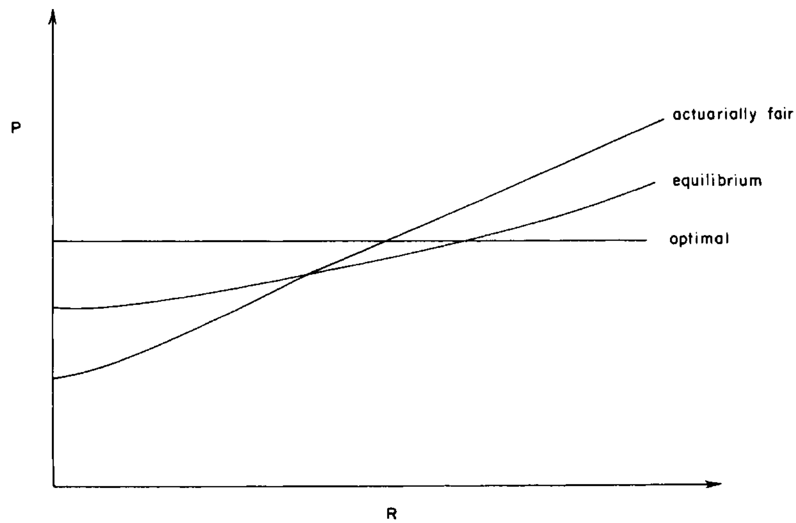

Fig. 10.5 Pensions as a function of retirement age.

The problem becomes more complicated if, as might be expected, disutility (perhaps due to illness) is negatively correlated with life expectancy. The expected cost of offering a pension plan will be greater than in the case where retirement decisions and life expectancy are unrelated. This phenomenon is referred to as adverse correlation. The effect of adverse correlation may be of actuarial advantage to late-retiring cohorts.

An interesting area of speculation concerns the role played by the distribution of disability dates. In an optimal plan, the relative rewards to individuals retiring at ages 65 and 70 may depend on the sizes of those two groups. Yet the sizes of the two groups could hardly have anything to do with actuarial fairness. Workers who are disabled early may have relatively lower pensions to counter the problem of adverse selection. If these early retirees received an actuarially fair pension, large numbers of relatively healthy workers might prefer early retirement, an outcome which is not feasible. The distortion away from the first-best solution is greater for young retirees than for old retirees; favorable early retirement options create moral hazard problems for a relatively large number of workers while favorable late retirement options have little adverse effect on workers who had to retire early.

10.4 Merged Pensions and Strict Pareto Inferiority

A major theme of informational economics is that inefficiencies are generated when individuals with differing characteristics are treated alike. The equal treatment may result from a regulatory requirement or an inability to distinguish between workers' characteristics. Despite the overall inefficiency, one group usually benefits at the expense of the other. High-risk drivers benefit when they are lumped in the pool with low-risk drivers. Under ERISA, all companies are charged the same premium for insurance; this helps the poorly funded pension plan at the expense of the well funded.

The effects of pooling are of considerable importance in retirement plans. The government is becoming increasingly involved in legislation in this area. Raising the permissible mandatory retirement age creates a problem of separating populations that did not previously need to be separated. Presently, there is serious discussion about eliminating all mandatory retirement ages. Legislation that requires pensions to be constant across a firm becomes more significant with the growth of conglomerates. When two disparate firms merge, there is a tendency to integrate towards the more generous pension plan, in part because of contractual obligation. The issue is perhaps most salient with regard to discussion of altering pension programs, which at present frequently differ by sex.[17] As workers have very different needs, desires, and life expectancies, there is no single neat solution to the social security problem. There will be an inevitable tug-of-war between the white-collar workers who want to work into their later years and the blue-collar workers who want to retire as soon as possible. Interestingly, both groups might be better off if each were able to have a plan of its own.

What is the cost of combining two pension plans? This question cannot be answered in general, for it depends on how the workers vary. To simplify, consider a company with two groups of workers, A and B. One Pareto optimal outcome will have the preferred pension for group A; call it $P^*(A)$, and likewise for B and $P^*(B)$. The constraint is that each of these contracts must be chosen from among those that break even. Assume for further simplification that there is no variability within groups.

There is no loss in combining the plans if the A's prefer $P^*(A)$ to $P^*(B)$ and the B's prefer $P^*(B)$. Our experience with insurance models suggests that this may not be the case. If, as U.S. law requires, the two groups are merged under a common pension offering both plans, the A's might choose $P^*(B)$. This leads to a problem since $P^*(B)$ is not achievable when used by both A and B type workers. The general solution under such circumstances is to alter $P^*(B)$, making it less attractive to the A's until they just choose their own plan (see Nichols and Zeckhauser 1982). At the same time, the plan for the A's need not be self-supporting; there can be a cross-subsidy to the plan intended to attract the A's. If so, we may get a pair made up of a somewhat distorted $P^*(B)$ that makes money and a subsidized $P^*(A)$.[18]

The pension problem is considerably more complex than the standard area of application of information economics. It leads to qualitatively new phenomena. It is quite possible that the optimal plan for A's in isolation will be chosen by the B's and that the preferred plan for the B's in isolation will be chosen by A's. If so, both groups may suffer losses from the merger of their pension funds. This is demonstrated below.

Two companies with different type of employees have each worked out their own optimal pension plan. All contributions have already been made and the per capita funds in the two plans are equal at 3/2 per worker. There are two periods remaining. The employees of the two firms have the common expected utility function

$$EU = U(C_1) + p_s U(C_2) - D^i(L), \tag{48}$$

where C_i = consumption in periods i, $i = (1, 2)$, p_s = the probability of survival in the second period, L = the first-period labor supply, either 0 if retired or 1 if working, $D^i(L)$ = the disutility of work, $D^i(0) = 0$, $i = (1, 2)$. The functional form of the utility function is the same for all employees. It is concave, hence risk averse, and is illustrated in figure 10.6.

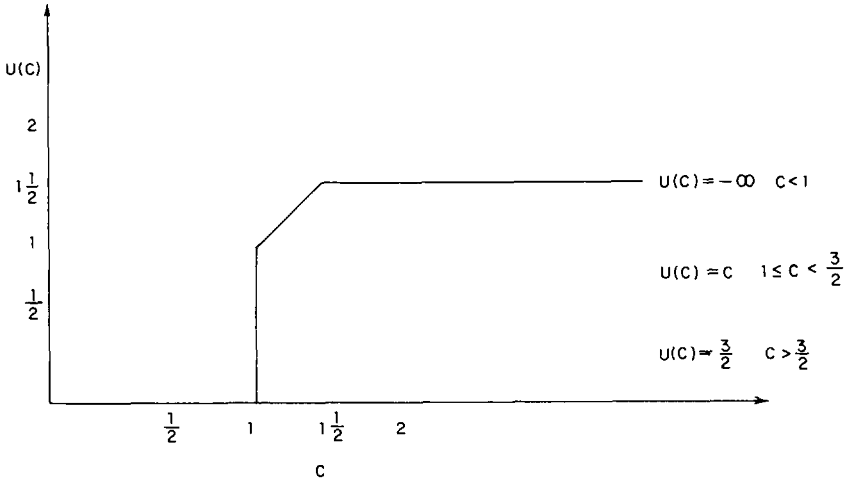

Fig. 10.6

(49)
$$U(C) = -\infty, C < 1,$$
$$U(C) = C, \ 1 \le C < \frac{3}{2}$$
$$U(C) = \frac{3}{2} C > \frac{3}{2}.$$

The employees of company A have the following characteristics

(50)
$$p_s = \frac{1}{2}; D^A(1) = 3/5; \ \textit{marginal product} = 0.$$

Since their marginal product is zero, it is optimal for them to retire immediately. With their accumulated pension savings of 1½ units per capita, the optimal plan yields

(51)
$$C_1^A = 1; \ C_2^A = 1$$

(52)
$$EU^A = U(1) + \frac{1}{2} U(1) - D(0) = 1\frac{1}{2}.$$

Note that this is also actuarially feasible, since the expected total pension $= C_1^A + p_s C_2^A = 1½$, which is the original accumulated pension savings.

The employees of company B have the characteristics

(53)
$$p_s = 1; \ D^B(1) = 5/4; \ \textit{marginal product} = 3/2.$$

Here, the optimal strategy for a B employee is to work in the first period, retire in the second. He receives

(54)
$$C_1^B = 3/2; \ C_2^B = 3/2$$

(55) $$EU^B = U(3/2) + U(3/2) - D^B(1) = 1\frac{3}{4}.$$

This consumption pattern is also actuarially supportable, since expected total consumption $= C_1^B + p_s C_2^B = 3 =$ total pension savings + marginal product.

Now assume that the two pension plans are merged. Who will benefit and who will lose? First, notice that if offered the opportunity, any *A* would prefer *B*'s pension plan to his own. That is, he would work in the first period if by doing so he could earn a consumption stream of 3/2, 3/2. This yields

(56) $$EU^A = U(3/2) + \frac{1}{2}U(3/2) - D^A(1) = 1\frac{13}{20}.$$

And, somewhat surprisingly, *B* would prefer *A*'s plan. *B* would retire in the first period with a consumption stream of 1,1.

(57) $$EU^B = U(1) + U(1) = 2.$$

The comparisons are as follows:

	Expected Utility	
	A	*B*
A's plan	$1\frac{1}{2}$	2
B's plan	$1\frac{13}{20}$	$1\frac{3}{4}$

This paradoxical situation arises only when *A*'s pension program is not supportable if *B*'s comprise the population, and *B*'s is not supportable if *A*'s comprise the population. The difficulty is that any program that gets the productive *B*'s to work will also induce the nonproductive *A*'s to work; if the low life expectancy *A*'s retire, the *B*'s will also retire.

The choice for the merged pension plan is to have both groups retire or to have both groups work in the first period. Given this constraint, there exists no pension plan that breaks even and offers either group a situation as favorable as the one with which it started.

Let there be equal numbers of *A*'s and *B*'s, normalized to one for each group. Assume that both retire early. There is a total of three units of pension savings. Given that $U(C) = -\infty$ for $C < 1$, we must have $C_1 = C_2 = 1$. But this consumption program will take 1½ units of consumption for *A* and 2 for *B*. It is infeasible. Any plan that induces both individuals to retire in the first period must offer each of them a lower expected utility than he had with his initial plan.

Consider now plans that induce both A and B to work for one period. The total resources available will be 3 of pension savings plus 1½ from B's production, which equals 4½. The first-best outcome for B is to receive a consumption stream of 3/2, 3/2, since additional consumption offers no utility. But this consumption stream is not supportable for the combined groups, because A would require 3/2 in the first period and his probability of survival, 1/2, times 3/2 in the second, totaling 2¼. Adding to B's requirement of 3, the required resources are 5¼, but only 4½ is available.

The best supportable consumption stream for A is 3/2, 1. (Both A and B get 3/2 in the first period. A gets an expectation of 1/2 in the second period while B gets 1). But since A's disutility of work is 3/5, he too must lose from a merged pension plan where both work.

The conclusion is that all pension plans for the merged group leave both parties strictly worse off than they were in isolation. This demonstration was facilitated by having a utility function with vertical and horizontal segments, but the possibility for a counterexample is general and could be created for a utility function with continuous curvature.

A policy conclusion emerges from this demonstration. Considerations of risk spreading and equity have been used as arguments for homogenizing pension programs across different types of workers and firms. Sometimes the restrictions are specific: the pension plans must be the same. Other times regulatory impositions, such as some of those associated with ERISA, encourage conformity. This example shows that it is important to ask whether the imposition of "equal treatment" in retirement programs may hurt many or most if not all.[19]

10.5 Conclusions

Labor markets are quite different from the markets studied in introductory textbooks. The products sold are heterogeneous, substantial uncertainties are compounded by asymmetric information flow, and interdependencies in "sales" may extend across several periods because of such matters as training and proprietary information. Given such complexities, secure contracting possibilities and flexible reward schedules (i.e., prices that depend on outcome,) would be particularly desirable. However, rules and standards against indenturing limit possibilities in the first area. The second is constrained by a variety of institutions that limit wage flexibility, and by quite considerable risk aversion on the part of sellers in the market (i.e., the workers).

Pensions are a powerful instrument for compensating labor. They may be able to facilitate labor market operations. They offer several significant advantages: awarded late in life, pensions can be based on extensive performance; they are granted in periods to which workers might wish to reallocate resources on their own; and they offer many degrees of freedom in a reward structure that can depend on such factors as salary, age, and years of service.

The purpose of this analysis was to provide a framework for examining the effects of pensions on retirement decisions. At least since the 1920s, the advantages of pensions as a mechanism for securing the retirement of workers whose productivity is falling have been understood and applied. This feature is likely to become more important with the passage of time for at least three reasons: (1) individuals are living longer; (2) as the workplace becomes more technologically sophisticated, the dangers of technological obsolescence of older workers become more pronounced; and (3) both overtly, as through legislation, and implicitly, society is increasing its "protection" of workers who might not choose to retire. The recent congressional decision to raise the mandatory retirement age of private sector workers to 70 has now been extended to state and local government employees. Age discrimination suits associated with layoffs and dismissals have become relatively commonplace. There is talk, indeed, about abolishing mandatory retirement provisions altogether.

We have begun by focusing on the factors that induce individuals to retire. The first model incorporates the qualitative features of most defined benefit pension plans in the United States and inquires how such plans can be employed to induce optimal retirement decisions. Pensions are viewed as a form of forced saving whose purpose is to enable the worker to "commit himself" by making it in his own self-interest to retire at an appropriate age.

The remaining models examine the use of pensions in populations that are heterogeneous with respect to such features as disutility of work or expected life span. Given heterogeneity, a major policy concern is whether pensions are actuarially fair to different groups, retirement cohorts, and so on. An optimal pension plan cannot be actuarially "more than fair," in the sense that someone who retires later must impose a smaller cost on the pension pool than he would if he retired earlier. However, people who retire later are likely on average to live longer. Under most common pension plans, late retirees impose a greater cost on the pension fund than those retiring earlier.

In a first-best world, a separate pension plan would be designed for each group of workers. (Conceivably there would be lump-sum transfers among plans. They need only break even as a whole.) However, government-mandated retirement programs, legislation regulating private pension programs, and the forced common form of pension programs within single firms are powerful forces for homogenization. Such homogenization is shown to work to the possible detriment of workers as a whole.

Pensions are a workhorse compensation mechanism, meeting a variety of objectives in labor market operations and dealing with a range of imperfections and contracting difficulties. This analysis has focused on the ideal use of pensions to facilitate appropriate retirement decisions. The challenge both for conceptual work and for policy is to design pensions that blend the objectives of attracting, sorting, motivating, and retaining workers with the need to induce reasonable retirement choices.

Appendix
Pension Benefits and Retirement Age

A survey of noncontributory pension plans from 75 United States companies revealed several patterns about the general structure of private pensions in this country. These companies, which had businesses ranging from communications and insurance to manufacturing and mining, all had prescribed minimum requirements at least for the age of retirement and often for the years of service as well. Benefits were calculated on a monthly basis in each case, the most common formulas multiplying the years of service by a set dollar figure or by a percentage of salary. While most companies utilized the standard retirement age of 65 as the minimum requirement, several allowed retirement at 62 or even 60. Often the workers were given an option of retiring, for instance, at 65 with 10 years of service or at 60 with 30 years of service. It thus becomes difficult to classify the retirement ages encouraged by pensions.

The benefit formulas ranged from the exceedingly simple to the bewilderingly complex. Many companies offer the employees the option of choosing among several formulas that take into account, among other things, years of service, average monthly pay over the last five years, and future social security benefits. Others simply multiplied the years of service the worker had put in with the firm times a set dollar amount. Most of them, whether simple or complex, incorporate either a set dollar amount or a percentage of salary, multiplied times the years of service. A typical plan might offer the worker a monthly benefit of his years of service multiplied times either $15 or 1.5% of final monthly earnings. Extra percentages were offered by a few companies to workers with 30 or more years of service. The number of variables in the benefits formula make it difficult to directly compare benefit payments across companies.

All of the pension plans incorporate an early retirement option that allows the workers to retire before the prescribed minimum age, while imposing a penalty if they want payment before they reach the normal retirement age. These penalties usually deduct a set percentage of the pension for each year that the employee retired early. Many companies take into account other factors such as years of service; the early retirement penalty thus varies across workers. Forty-six of the companies surveyed utilized a uniform percentage reduction system. These penalties ranged from 3% to 7% for every year under the prescribed age; most were between 4% and 5%. If an employee retired at age 60 when the minimum retirement age was 65 and a 5% per year penalty was imposed, that worker would receive 75% of the benefit he would normally receive. In a few cases, the penalties are scaled down when early retirement occurs within three years of the normal retirement age.

Notes

1. The union attempted to counter these employer-sponsored plans by instituting independent pension plans of their own. While the firms concentrated on providing future old age benefits, the unions initially offered only general benefit programs; immediate benefits for sickness, disability, death, and strikes appealed more to the younger workers they were trying to attract. They gradually expanded the scope of program benefits and soon focused primarily on old age payments. By 1928, about 40% of union members belonged to national unions offering one form or another of old age benefits. These pensions were funded by assessments on union members, and they became increasingly burdensome as the number of older workers increased. Raising union dues became more difficult, especially with the advent of the Depression. Other demands on the unions' treasuries, combined with the nation's financial chaos, resulted in the almost complete collapse of all union welfare plans by the early 1930s. After social security was adopted in 1935 only a handful of union plans survived, and it was not until the Second World War that union interest in pensions revived.

2. At present, this is done through the Supplemental Security Income program.

3. Pensions can also provide an effective wage adjustment if they are actuarially unfair to workers who postpone retirement (Lazear, in this vol.). However, as discussed later, pensions that appear unfair may only be a reflection of the heterogeneous life spans in the population.

4. As discussed earlier, there are tax advantages for savings in pensions. These deferred taxes result, all other things equal, in an effectively lower discount rate for savings held by the firm. This provides an additional argument for pensions.

5. Larger pensions increase workers' forced savings. To raise savings in equilibrium, we must assume that workers cannot fully counter this effect by borrowing against their pension to restore current consumption. We assume that there is no borrowing act of this sort.

6. Bulow (1981) demonstrates that workers who retire early should take advantage of the early retirement option rather than wait until the normal retirement age before collecting benefits.

7. If inflation were recognized in the design of the pension plan, and if it were constant, all calculations using real dollars would pertain. Given that most pension plans deal in minimal dollars, variable inflation rates impose risk costs on the worker unless inflation indexation is perfect and immediate.

8. One theoretical solution to this problem would be to stop paying pensions after 20 years of retirement. Indeed, some plans allow the workers to take the pension's actuarial value in a lump sum payment. If this were required rather than an option, workers with longer life spans would have less incentive to take advantage of early retirement provisions.

9. This problem is complicated if, as is common, the pension plan gives a surviving spouse a substantial fraction of the worker's benefits. The problem will be mitigated if variability in expected benefits is less once spouse's benefits are included. We do not consider survivors' benefit in this analysis, leaving for the future such interesting questions as the correlation among life expectancies of worker and his spouse (i.e., do young women seek as husbands older workers with high pensions offering survivor clauses?) or the extent to which workers take into account when making retirement decisions the welfare of their spouse after their own death.

10. In conventional plans this may be some average of final-year wages. In pattern plans, the function would be a constant.

11. In theory, the corporation could make a contract for the way it treats the work force as a whole. For example, it could commit itself to have no more than 25% early retirements. An alternative means of allowing for differential treatment, yet providing adherence to earlier contracts, is to offer to pay someone (not necessarily the worker) an amount that depends on

the action that is taken relative to the worker. This latter case simply creates a mechanism to make the contract self-enforcing.

12. Cooper (1983) nicely categorizes some general properties of the optimal second-best solutions in problems with adverse selection or moral hazard when the agents' preferences obey SCP.

13. Disabilities do shorten life spans. Models of pensions with heterogeneous life expectancies are considered in the previous section.

14. Workers who are disabled before the minimum retirement date stop working but do not start collecting pensions until age R_0.

15. For small values of R, $\partial B/\partial R$ is negative. To see that there is a unique solution to $\partial B/\partial R^* = 0$, observe from eq. (45′) that at any solution to $\partial B/\partial R = 0$, the second derivative $\partial^2 B/\partial R^{*2} = -U''\,[P(R^*)]P'(R^*)(T-R)/U'\,[P(R)]^2$, which is positive.

16. The second derivative of the actuarially fair curve is $P'' = 2P'/(T-R)$ while the curve that keeps workers indifferent about retirement age has a larger second derivative, $P'' = 2P'/(T-R) - U''P'^2/U'$. Since the curves are tangent at R^*, the actuarially fair curve must lie below the indifference curve for $R < R^*$.

17. In 1983, the Supreme Court in *Arizona Governing Committee v. Norris* held that for tax-deferred compensation plans, the monthly retirement benefits cannot be lower to women who made the same contributions as men.

18. Such a subsidy is not possible in a competitive market in which the A's and B's can be distinguished. A firm could make positive profits by offering only the $P^*(B)$ plan and not hiring and A type workers.

19. Negative outcomes of this type are particularly likely if other aspects of the labor packages get adjusted to take account of retirement packages. If so, there might be a presumption that any attempt to eliminate distinctions in retirement programs that promote inefficiency would work to the detriment of all.

References

Bulow, J. I. 1981. Early retirement pension benefits. NBER Working Paper 654.

Cooper, R. 1983. Towards a solution to the general problem of self-selection. Yale University Working Paper.

Diamond, P., and Mirrlees, J. 1978. A model of social insurance with variable retirement. *Journal of Public Economics* (December).

Feldstein, M. 1982. The optimal level of social security benefits. NBER Working Paper 970.

Greenough, W. C., and King, F. P. 1976. *Pension plans and public policy*. New York: Columbia University Press.

Hausman, J., and Wise, D. 1985. Retirement and subsequent work. In this volume.

Haynes, S., McMichael, A.; and Hyroler, H. 1979. Survival after early and normal retirement. *Journal of Gerontology*.

Lazear, E. P. 1985. Pension benefit formulas and worker productivity. In this volume.

Nichols, A., and Zeckhauser, R. 1982. Targeting transfers through restrictions on recipients. *American Economic Review* (May).

Woodbury, S. A. 1983. Substitution between wage and nonwage benefits. *American Economic Review* 73 (March): 166–82.