LABOR TURNOVER, WAGE STRUCTURES, AND MORAL HAZARD: THE INEFFICIENCY OF COMPETITIVE MARKETS

Richard J. Arnott
Joseph E. Stiglitz

Econometric Research Program
Research Memorandum No. 289

September 1981

Richard Arnott
Queen's University
Kingston, Ontario
Canada K7L 3N6

Joseph Stiglitz
Princeton University
Princeton, NJ 08544
U.S.A.

ABSTRACT

"Labor Turnover, Wage Structures, and Moral Hazard:

The Inefficiency of Competitive Markets"

by

Richard J. Arnott

Joseph E. Stiglitz

This paper is related to three bodies of literature -- those on moral hazard, multi-period employment contracts in the Becker tradition, and implicit contract theory.

A multi-period, general equilibrium model of the labor market is developed in which risk-averse workers are faced with job-related uncertainty and labor turnover is costly. If a worker is unlucky and suffers a bad job match, he quits and joins another firm, hoping that he will like its work environment more. We assume that the quality of a job match is unobservable; as a result, insurance markets are incomplete. We assume furthermore that the firm, rather than some other agent, provides partial, implicit insurance against job dissatisfaction, and does this by paying workers more than their marginal product in their early years with the firm, and less subsequently. Such implicit insurance is a form of quit insurance. Since the probabilities of the insured-against events (the quit rates over time) are affected by the amount of such insurance provided, the firm's implicit insurance is characterized by moral hazard. We show that this moral hazard gives rise to (constrained) inefficiency and indicate why.

LABOR TURNOVER, WAGE STRUCTURES, AND MORAL HAZARD:

THE INEFFICIENCY OF COMPETITIVE MARKETS*

I. Introduction

One of the most significant developments in micro-economic theory during the last two decades has been the exploration of economies in which information is imperfect, or more properly asymmetric. It is by now well-known that in such economies, insurance markets are typically incomplete and are characterized by moral hazard and adverse selection problems. What is not so widely appreciated is that this incompleteness of insurance markets significantly affects other markets in the economy. Institutions develop in these other markets to provide implicit insurance.[1,2] A question of central importance is whether these accommodating institutions are efficient, conditional on the information technology. This paper addresses this issue by focusing on a feature of the labor market, the employment contract.

In the traditional analysis of the structure of competitive multi-period employment contracts, it is assumed that workers, even though they face job-related uncertainties, choose to work for the firm offering the highest discounted present value of expected future remuneration (income plus income in kind). This assumption requires either that workers be risk-neutral or else that capital markets be perfect and insurance markets complete,[3] and implies that workers pay the full cost associated with their being hired and are paid the net (of training costs) value of their marginal product at each point in time. In these circumstances, the equilibrium set of employment contracts is efficient. Realistically, however, workers are risk-averse and, because of informational asymmetries, face incomplete insurance markets and imperfect capital markets. What matters to the worker, then, is expected utility rather than the PDV of future remuneration. In this situation, firms may have an incentive to provide implicit insurance against job-related contingencies in their employment contracts.[4] One way in which they can provide such insurance is by paying lucky workers less than the value of their marginal products, and unlucky workers more.[5] Thus, employment contracts may have qualitatively different characteristics in economies with informational asymmetries.

This paper addresses the following questions: What are the characteristics of competitive multi-period employment contracts when information is imperfect? How do these characteristics differ from those of employment contracts in economies with perfect information? Is the former class of contracts efficient, conditional on the information available? If not, what are the sources and consequences of the market failures, and what government intervention would eliminate or minimize the excess burden associated with them? To address these questions, we consider a very simple economy with a single kind of uncertainty. Nevertheless, the economic principles derived can be usefully applied in the examination of a wide range of social insurance issues.

The emphasis of this paper is on the <u>efficiency</u> of competitive employment contracts with imperfect information. The descriptive, comparative static properties of such contracts have been treated in a companion paper, Arnott [1980]. The market failures we identify in this paper are generic and are treated more generally in Arnott and Stiglitz [1981 a,b].[6]

When we started this paper, we conjectured that in economies with informational asymmetries, the competitive equilibrium set of multi-period employment contracts would be inefficient. Our reasoning went as follows: Imagine that the first-best[6a] set of employment contracts (in which workers are paid the value of their net marginal products and pay for the costs of their being hired) is offered in an economy in which no job-related insurance is available. Relative to the first-best optimum, there is a deadweight loss associated with the unavailability of insurance. This deadweight loss can be reduced if firms provide implicit insurance in their contracts. But with informational asymmetries, moral hazard problems arise when insurance is available. For instance, if the insurance takes the form of the firm's paying part of the loss in income incurred by a worker if he quits, then the provision of insurance results in excessive (more than in the first-best optimum) labor turnover, and there is a deadweight loss associated with this. In the constrained optimum, the sum of these two deadweight losses is minimized; the marginal benefit of providing more insurance (the value of the reduction in the risk the worker faces) equals the marginal cost (the cost of the distortions induced by the provision of insurance). Now consider two firms, the second of which hires the workers that quit the first. The second firm, in deciding on the contract to offer, will ignore the effect its contract has on the proportion of workers who quit the first firm. This would not matter in a first-best economy. But here workers who quit are being subsidized. Thus, the second firm fails to take into account that its actions affect the profitability of the first firm. The phenomenon entails a form of externality, which causes the competitive equilibrium to differ from the constrained optimum.

It turns out that the above line of reasoning is incomplete. Market failure does occur but for reasons that are more complex and subtle than the above argument suggests. The types of market failures which arise depend on the characteristics of capital markets. Furthermore, when capital markets are perfect, market failure occurs only in economies with three or more periods.

The above argument does, however, correctly identify the central issue - whether, in this context, competitive markets will efficiently handle the trade-off between risk-bearing and incentives.

We employ the same model, with slight modifications, throughout the paper. A single numéraire commodity is produced according to a single technology which exhibits constant returns to scale to the sole factor, labor. The risk-averse workers are equally productive and have the same utility function with respect to income and job satisfaction. They differ, however, in what they like and dislike about a particular job; some enjoy rotating shifts, others do not; some enjoy Musak on-the-job, others do not; and so on. Firms are competitive and risk-neutral. They offer different work environments; some have rotating shifts, others do not; some have Musak, others do not; etc. At the time a worker joins a firm, he does not know what the firm's work environment is like, nor can the firm judge *ex ante* which workers will enjoy the particular work environment it offers. Thus, when joining a firm, a worker is uncertain concerning how much he will enjoy his work environment (or, synonymously, what his level of job satisfaction will be with the firm, or what the quality of the job match will be). This is the only kind of uncertainty in the economy. While workers' tastes vis-à-vis the work environment differ, they are symmetric in the sense that the probability distribution of a worker's job satisfaction when he starts employment with any firm is independent of both the firm and the worker.[7,8] A worker quits a firm if doing so increases his expected utility. The replacement of this worker by another entails a fixed cost to the firm.

The qualitative characteristics of the set of equilibrium employment contracts depends on the information technology. We assume that the firm knows a worker's employment history[9] at the time he joins the firm, but cannot ascertain his job satisfaction while he is with the firm. A worker, meanwhile, has full information on the contracts offered by all firms, but does not know, at the time he joins a firm, how much he will enjoy his job. Finally, the firm can acquire insurance-relevant information concerning its workers at a substantially lower cost than can any agent external to the firm.

Before presenting the model formally, we comment on some of its characteristics and on the role played by some of the assumptions. We have chosen the assumptions so as to provide as simple as possible a general equilibrium model which contains multi-period employment contracts, labor mobility, and incomplete insurance markets. The qualitative results to be derived will clearly generalize to more complex and realistic economies.

1. The assumptions <u>imply</u> what kind of insurance is provided and by whom, in contrast to most other papers which instead assume these characteristics.

The firm or any other insurer is unable to observe a worker's job satisfaction, but can observe whether or not he quits. Suppose that full insurance were provided against job (dis)satisfaction. When asked about his level of job satisfaction, the worker would have an incentive to lie; specifically, he would state that his job satisfaction was the lowest possible consistent with those aspects of his behavior observable by the insurer, specifically whether he quits or stays. In fact, there is no efficient method of information revelation. Full insurance is not therefore issued against job dissatisfaction and insurance markets are incomplete. This stems from an asymmetry in information -- the worker knows his job satisfaction, but the insurer can neither observe it nor ascertain it costlessly.

The question now arises as to what agent will provide insurance and in what form. We have assumed that firms have an informational advantage over external agents in the provision of insurance, and will therefore provide the insurance.[10] This assumption is easily justified for a more complex economy than the one we consider. For instance, suppose the worker is paid different amounts depending on whether his separation from a firm is due to his having quit or having been fired. The firm can monitor this better than an external agent. In fact, if an external agent provides such insurance, the firm and the worker have an incentive to collude to obtain the higher payout for the worker. The assumption is more difficult to justify in the context of the economy of this paper, however. If we were instead to assume that external agents provided insurance, market failures analogous to those identified in the paper would arise.[11] There would be an additional source of market failure if the firm and the external agent were to provide insurance simultaneously - each agent would ignore that the amount of insurance it provided affected the profitability of the other's contract.[12]

It is by now well-known that with informational asymmetries, an insurance contract specifies both the premium and payout, i.e. not only the price of insurance but also the quantity that may be bought. In the context of this paper, in which a worker's employment history, but not his job satisfaction, is observable, the richest form of insurance contract that could be offered would have the following characteristics: For the first period, the contract would specify the insurance premium, and the payout in the event the worker quit at the end of the period. At the end of the first period, all stayers would be offered a continuum of premium-payout pairs for the second-period across which they would self-select on the basis of their realized first-period job satisfaction.[13] All quitters would be offered a different continuum of premium-payout pairs by other prospective employers. And each of these premium-payout pairs would specify two (one for quitters, one for stayers) continua of premium-payout pairs between which a worker could choose at the end of the second period and applying to the third period. And so on. Since such contracts specify the premia and payouts contingent on employment history and not on job satisfaction, they are providing quit insurance. We rule out such a rich form of quit insurance contract because of the prohibitive cost of administering it. We assume instead that an insurance contract specifies a unique premium-payout pair for each employment history -- a simpler form of quit insurance. If we were to treat the more complex contracts, we would obtain the same sources of market failure as with the simpler form of contract dealt with here.

The quit insurance can manifest itself in a number of institutional forms. It can involve explicit premia and payouts. It can also be provided implicitly. Normally, the luckier a worker is (the higher the quality of his job match), the longer will he stay with the firm. Thus, implicit insurance against job dissatisfaction can be provided by paying those who quit soon after joining the firm more than their marginal product and those who stay with the firm a long time less. This can be achieved by setting a wage schedule as a function of length of tenure, with the wage above the marginal product during the early years and below the marginal product subsequently. Alternatively, all workers could be paid less than their marginal products, and the surplus could be used to finance a severance pay fund.

2. An important characteristic of our economy is that a firm must break even on the insurance policy it offers each group of workers who join the firm with a particular employment history. Cross-subsidization is not possible. Suppose it were otherwise and that a firm could use the profit from the insurance provided to a lucky group (e.g., fifty-year-olds who had only had one previous job) to subsidize the insurance provided to an unlucky group (e.g., thirty-year-olds who had already had five previous jobs). If another firm were to offer a slightly more attractive policy to the former group and a less attractive one to the latter, it would attract only the former group, thereby making a profit.

3. For the model to be interesting, we must ensure that there is some inter-firm job mobility. If each worker who quit his job were to take another job in the same firm, then the externalities we identify would be internalized and there would be no market failure. There are several ways to do this. We could assume that a number of the attributes of a job which affect a worker's job satisfaction are firm-specific rather than job-specific. Or we could assume that some job attributes are location-specific and that, because of administrative diseconomies of scale external to individual plants but internal to firms, each firm operates at only one location. In either case, when a worker changes jobs, he may decide to change firms.[14]

The externalities could also be internalized if different firms were to co-ordinate the setting of wages. We rule out this possibility by appealing to the costs of such co-ordination. With this assumption, to obtain market failure in our economy, all we require is that there be <u>some</u> mobility between firms. To simplify the analysis, however, we make the unnecessarily strong assumption that there are no internal labor markets. Thus, when a worker changes jobs, he changes firms.

In comment 2, we argued that a firm must break even on the insurance policy (and thus on the employment contract) it offers each group of workers who join the firm with a particular employment history. Combining this observation with the assumption of no internal labor markets leads to an analytically convenient definition of a firm, which we adopt, as a group of workers in the same job who had the same employment history when they started the job. Every such firm must break even.

4. Moral hazard is present whenever the provision of insurance affects the probability of the insured-against event. There is moral hazard in our economy since the insured-against event is quitting, and the probability of a

worker's quitting depends on the costs of doing so, which is influenced by the amount of insurance provided. Some of the implications of moral hazard are by now well-known. In particular, equilibrium contracts provide only partial insurance and specify not only the price but also the quantity of insurance to be purchased.[15] However, the welfare economics of moral hazard are not well-understood. In the existing literature, it is argued that moral hazard, by itself, does not upset the optimality (conditional on the information available) of competitive markets, but this has been demonstrated only for very simple economies with one source of uncertainty and one consumer good. We shall show contrarily that in our economy the fact that the provision of insurance is characterized by moral hazard does cause market failure. Thus, this paper may be viewed as examining, in an admittedly rather specific context, some aspects of the welfare economics of moral hazard.

The major source of market failure we identify is that associated with what we term seemingly unrelated events. We shall now give an example of this market failure and then show how it applies to the problem at hand. Suppose that there are two statistically independent sets of states of nature. One set is associated with the probability of one's housing burning down; the other, with having an automobile accident. Because of asymmetric information (i.e., the insurer can observe the outcome or event, but neither the underlying state of nature nor the insured's accident-prevention effort), insurance is provided against the two events rather than against the underlying states of nature and both forms of insurance are, as a result, characterized by moral hazard. Suppose, furthermore, that an individual purchases insurance against one accident from one agent and against the other accident from another. The provision of insurance against his housing burning down may cause the individual to become less risk-averse and consequently to drive less carefully, thereby increasing the probability of an automobile accident. In this case, the provider of fire insurance ignores in his calculations that by providing more fire insurance he affects the profitability of the contract offered by the provider of automobile accident insurance. Put alternatively, the private cost he perceives in providing an extra unit of insurance will in general differ from the social cost; as a result, he provides the wrong amount of fire insurance. The two agents fail to take into account this interdependence between the insurance contracts they offer.

In the problem at hand, the two statistically independent risks are the qualities of two job matches for the same worker, one corresponding to a job earlier in his career, the other to a later job. The interdependence comes about through saving. The firm that employs the worker early in his career fails to take into account that the contract it offers, by influencing the amount the worker saves, will in general affect the profitability of the firm that employs the worker later in his career. For example, if the former firm provides more implicit insurance in its employment contract, this may cause the worker to save more which may make him less risk-averse and therefore increase the probability that he quit the latter firm. Similarly, the latter firm fails to take into account that the contract it offers will in general affect the profitability of the former firm.

5. We assumed that firms are identical and that workers are identical in all relevant respects except those related to the quality of a job match in order to circumvent the adverse selection problems which would otherwise arise. These problems have been treated in Rothschild and Stiglitz [1976]. Abstracting from adverse selection not only simplifies an already complicated analysis, but also allows us to ascribe the market failures we identify to moral hazard.

It will be useful to summarize the main points made in the above comments. In the economy we have described, we define a firm to comprise a group of workers in the same job who had the same employment history when they started the job. Each firm, defined as such, must break even and in equilibrium will offer an employment contract that provides partial insurance against quitting. The insurance may manifest itself in a variety of institutional forms; one involves setting the wage of less senior workers above, and that of more senior workers below, the corresponding marginal product. The provision of insurance is characterized by moral hazard, but not by adverse selection. The presence of moral hazard typically results in market failure.

We now proceed with the analysis. Section II provides a simple introduction to the model by examining a two-period economy, in which competitive employment contracts are efficient when capital markets are perfect, but not otherwise. Unfortunately, these efficiency characteristics are particular to the two-period model. To obtain the qualitative efficiency characteristics of an n(>2) - period or continuous time economy, it is necessary to treat the three-period economy. This is done in Section III. Several sources of market failure are identified and discussed. Section IV presents concluding comments.

II. The Two-Period Economy

In this economy workers live and work for two equal-length periods. At the beginning of his career, a worker decides to join that firm whose employment contract yields him the highest expected utility over his career. (In equilibrium, all firms offer the same contract.) Upon starting work, he finds out how much he enjoys his job. This level of job satisfaction remains constant during the first period, but may change between the first and second periods according to some known stochastic process. The worker may not quit this firm until he has worked there one period. At the end of the first period, he decides whether or not to quit. He knows what his second period wage will be if he does and does not quit, and also the probability distributions of second-period job satisfaction in the two cases. He quits if doing so raises his expected utility. At the beginning of the second period, having made his decision, he finds out his second-period job satisfaction which remains constant over the period. He must remain with his second-period firm for the entire period.

This two-period economy has two types of firms (according to our definition) - those that hire workers when they have just joined the labor force, and those that hire workers at the beginning of the second period of their working lives, after they have quit their first-period job. We label the two firm types 1 and 2 respectively. A type 1 employment contract is characterized by a first and second period wage;[16] a type 2 contract by a single wage.

The following notation is employed:

w_i^j	the wage offered by a firm of type j during the i^{th} period of a worker's life
θ_i^j	the job satisfaction of a worker in a firm of type j during the i^{th} period
$f[\theta_i^j]$	the p.d.f. of θ_i^j, with $F[\theta_i^j]$ the c.d.f. (throughout the paper we use square brackets to enclose the arguments of a function)
$U[\{w_i^j\},\{\theta_i^j\}]$	utility function in mixed direct-indirect form (strictly concave)
q_i^j	the proportion of workers who join a firm of type j that quit at the end of period i
T^j	hiring or training costs for a firm of type j
m	marginal product of labor.

The notation is more complex than necessary for the two-period model, but has the advantage that it can be extended to the three-period model. To simplify some of the later mathematical arguments, we assume that $f[\theta_i^j]>0$ for $\theta_i^j \varepsilon(-\infty,\infty)$ and is continuous.

We first solve for the competitive equilibrium, and then, to investigate its efficiency properties, determine the planning optimum.

We have argued that firms as herein defined make zero profits in competitive equilibrium. Costs per worker for a type 2 firm are $w_2^2+T^2$, while the revenue attributable to a worker is m. Thus, zero profits for a type 2 firm imply

$$w_2^2 = m - T^2. \quad (1)$$

Type 2 firms therefore have no flexibility in the wages they offer. Type 1 firms, however, do. Since workers are, by assumption, familiar with all wage contracts offered, they will choose to work for those type 1 firms that offer contracts providing the highest level of expected utility. Since the economy is competitive, equilibrium type 1 contracts maximize a worker's expected utility, subject to yielding zero profits.

We may view the determination of w_1^1 and w_2^1 as the solution to a two-stage maximization problem. In the first stage, workers decide on the subset of levels of period 1 job satisfaction corresponding to which they would quit, and the complementary subset corresponding to that which they would not,[17] taking w_1^1, w_2^1, and w_2^2 as given. This yields expected utility and the quit rate, q_1^1, as functions of w_1^1, w_2^1, and w_2^2. In the second stage, a type 1 firm chooses w_1^1 and w_2^1 to maximize workers' expected utility, subject to its making zero profits, taking into account the dependence of the quit rate on w_1^1 and w_2^1.

A worker's utility if he does not quit may be written in mixed direct-indirect form as $\hat{U}[w_1^1, w_2^1, \theta_1^1, \theta_2^1]$, and if he does quit as $\check{U}[w_1^1, w_2^2, \theta_1^1, \theta_2^2]$. To simplify notation, we assume that a worker will quit if his first-period job satisfaction is below some critical level, $\bar{\theta}_1^1$, and will stay otherwise (see footnote 17). Then a worker's expected utility may be expressed as

$$\breve{\mathcal{EU}} = \int_{-\infty}^{\infty} \int_{-\infty}^{\bar{\theta}_1^1} \check{U}[w_1^1, w_2^2, \theta_1^1, \theta_2^2]\, dF[\theta_1^1]\, dF[\theta_2^2]$$

$$+ \int_{-\infty}^{\infty} \int_{\bar{\theta}_1^1}^{\infty} \hat{U}[w_1^1, w_2^1, \theta_1^1, \theta_2^1]\, dF[\theta_1^1]\, dF[\theta_2^1 \mid \theta_1^1]. \tag{2}$$

In the first stage of the maximization, the worker chooses $\bar{\theta}_1^1$ as a function of w_1^1, w_2^1, and w_2^2; i.e.,

$$\bar{\theta}_1^1 = \bar{\theta}_1^1[w_1^1, w_2^1, w_2^2]. \tag{3}$$

Since $q_1^1 = F[\bar{\theta}_1^1]$, we may write

$$q_1^1 = q_1^1[w_1^1, w_2^1, w_2^2]. \tag{4}$$

And substituting (3) into (2) gives

$$\breve{\mathcal{EU}} = \breve{EU}[w_1^1, w_2^1, w_2^2]. \tag{5}$$

(4) and (5) together characterize the solution to the first stage of the maximization problem.

In the second stage, a type 1 firm's problem is (using (1)):

$$\max_{w_1^1, w_2^1} \breve{EU}[w_1^1, w_2^1, m-T^2]$$

$$\text{s.t.} \quad \text{i)} \quad q_1^1 = q_1^1[w_1^1, w_2^1, m-T^2] \tag{6}$$

$$\text{ii)} \quad B[w_1^1, w_2^1] = 0,$$

where B[·] is a type 1 firm's budget constraint. To further simplify notation here and throughout the rest of the paper, we assume that the discount rate is zero, in which case the budget constraint is

$$(2 - q_1^1)m = T^1 + w_1^1 + (1-q_1^1)w_2^1. \tag{7}$$

The term on the left-hand side is expected output per worker hired, while that on the right-hand side is the corresponding expected cost. Where λ is the Lagrange multiplier on the budget constraint, the first-order conditions are[18]

$$w_1^1: \quad \frac{\partial \breve{EU}}{\partial w_1^1} - \lambda(1 - \frac{\partial q_1^1}{\partial w_1^1}(w_2^1 - m)) = 0, \text{ and} \tag{8a}$$

$$w_2^1: \quad \frac{\partial \breve{EU}}{\partial w_2^1} - \lambda((1-q_1^1) - \frac{\partial q_1^1}{\partial w_2^1}(w_2^1-m)) = 0. \tag{8b}$$

(8a) states that w_1^1 should be raised to the point where marginal benefit equals marginal cost. $\left(\frac{\partial \breve{EU}}{\partial w_1^1}\right)\frac{1}{\lambda}$ is the expected dollar benefit to the worker from raising w_1^1 by one unit. There is a direct unit cost per worker, and also an indirect cost of $-\frac{\partial q_1^1}{\partial w_1^1}(w_2^1-m)$, the nature of which we now examine briefly. We have argued that the implicit provision of insurance will typically[19] involve paying less senior workers above their net marginal products and more senior workers below theirs. There is therefore a cost to the firm associated with a worker's quitting. $-\frac{\partial q_1^1}{\partial w_1^1}(w_2^1 - m)$ is the expected increase per worker in this cost from the unit rise in w_1^1. (8b) has an analogous interpretation.

We now return to the matter of central concern, whether the competitive employment contracts in this two-period economy are efficient, conditional on the information available. To ascertain this, we solve for the p.o. (planning optimum), conditional on the information available, and compare it to the c.e. (competitive equilibrium). The planner could regulate turnover directly, but since he does not know worker job satisfaction, this would result in transferring some workers who liked their jobs and not transferring other workers who did not, which would be inefficient. The best he can do is to set the three wages, and have the workers themselves decide whether or not to quit.[19a] Thus, the planning problem, too, is a two-stage one. The first stage is exactly the same as before, with the solution being characterized by (4) and (5). In the second stage, the planner maximizes workers' expected utility, subject to an economy-wide resource constraint, taking into account, via (4), how the quit rate responds to changes in the wage structure. The economy-wide resource constraint is

$$2m = T^1 + q_1^1 T^2 + w_1^1 + (1-q_1^1)w_2^1 + q_1^1 w_2^2. \tag{9}$$

The first-order conditions of his maximization problem are

$$w_1^1: \ \frac{\partial \breve{EU}}{\partial w_1^1} - \phi\left(1 - \frac{\partial q_1^1}{\partial w_1^1}(w_2^1 - T^2 - w_2^2)\right) = 0, \tag{10a}$$

$$w_2^1: \quad \frac{\partial \breve{EU}}{\partial w_2^1} - \phi\left((1-q_1^1) - \frac{\partial q_1^1}{\partial w_2^1}(w_2^1-T^2-w_2^2)\right) = 0, \text{ and} \tag{10b}$$

$$w_2^2: \quad \frac{\partial \breve{EU}}{\partial w_2^2} - \phi\left(q_1^1 - \frac{\partial q_1^1}{\partial w_2^2}(w_2^1-T^2-w_2^2)\right) = 0, \tag{10c}$$

where ϕ is the Lagrange multiplier on the resource constraint. These equations, which have an interpretation similar to that provided for (8a) and (8b), are to be compared with (1), (8a), and (8b), which together characterize the corresponding c.e.

Whether or not the planning optimum and the competitive equilibrium coincide depends on the nature of capital markets. In subsequent work, we plan to develop a typology to characterize capital markets. For our present purposes, however, it will suffice to define perfect capital markets to obtain when and only when any worker may borrow or lend as much as he wishes at the single market rate of interest. Whenever the condition is not satisfied, we say that capital markets are imperfect.

If capital markets are perfect,[20] then a worker is concerned with the sum of wages and not with the timing. Similarly, a worker bases his quit decision on the sum of wages he receives if he does and does not quit. Thus

$$\frac{\partial q_1^1}{\partial w_1^1} = \frac{\partial q_1^1}{\partial w_2^1} + \frac{\partial q_1^1}{\partial w_2^2} \quad . \tag{11a}$$

Also, from (2),

$$\frac{\partial \breve{EU}}{\partial w_1^1} = \frac{\partial \breve{EU}}{\partial w_2^1} + \frac{\partial \breve{EU}}{\partial w_2^2} \quad . \tag{11b}$$

Substituting these results into (10a)-(10c) reveals that, when capital markets are perfect, only two of the three equations are linearly independent. One may therefore set one of the wages arbitrarily. When w_2^2 is set equal to $m-T^2$, the equations characterizing the p.o. are the same as those characterizing the c.e. Thus, with perfect capital markets, the c.e. employment contracts in a two-period economy are efficient.[21]

The competitive economy has one more constraint than the planner - two zero-profit constraints instead of an economy-wide resource constraint. But with perfect capital markets, the worker is concerned only with the discounted incomes of a quitter $(w_1^1+w_2^2)$ and a stayer $(w_1^1+w_2^1)$. The constraint that $w_2^2=m-T^2$ is ineffective, since w_1^1 and w_2^1 can be adjusted so that any pair of discounted incomes for quitters and stayers attainable by the planner is also attainable in the competitive economy. Furthermore, because the type 2 firm wage is fixed by its budget constraint, such firms cannot affect the profitability of type 1 firms. Similarly, the contract offered by type 1 firms has no effect on the profitability of type 2 firms. Thus, there is no uninternalized externality of the type discussed earlier.

If capital markets are imperfect, however, the additional constraint imposed by the competitive economy is usually effective. Since the worker is now typically concerned with the timing of his wages, not just with the discounted value of his wage stream, it is no longer generally true that $\frac{\partial \breve{EU}}{\partial w_1^1} = \frac{\partial \breve{EU}}{\partial w_2^1} + \frac{\partial \breve{EU}}{\partial w_2^2}$ or $\frac{\partial q_1^1}{\partial w_1^1} = \frac{\partial q_1^1}{\partial w_2^1} + \frac{\partial q_1^1}{\partial w_2^2}$. In this case, there is a determinate optimal w_2^2 which most likely differs from $m-T^2$. To correct for this market failure, the planner could tax or subsidize a type 2 firm to bring its wage to the optimal level, financing the tax from, or redistributing the subsidy to, type 1 firms. Type 1 firms would then choose the optimal w_1^1 and w_2^1. We term this failure the cross-subsidization market failure, since the planner is able to cross-subsidize across firm types to achieve the optimum, whereas the competitive economy is not.[21a]

The question which naturally presents itself is: Should one expect capital markets to be perfect or imperfect in this economy? The literature provides little assistance in answering this question, since the nature of capital markets is almost invariably assumed rather than derived. We intend to develop a model in which the nature of capital markets is endogenous in a subsequent paper. In the present context, however, only a brief digression is meritted.

With perfect capital markets as we have defined them, a worker can borrow an infinite amount at the beginning of the first period. Should he default, the maximum penalty the bank can apply is to confiscate his second-period wage, plus any savings he has. Thus, perfect capital markets will not arise if, with them, even one worker would choose to default. This would occur if there were a single worker who would prefer infinite consumption in the first period and zero in the second to a finite level of consumption in both periods. In our stylized economy, both perfect and imperfect capital markets are possible.[21b] Which would occur depends on workers' tastes etc.

In conclusion, we have established that in a two-period version of our economy with perfect capital markets, the competitive employment contracts are efficient. With imperfect capital markets, however, they are generally not. Competitive equilibrium in this case suffers from the cross-subsidization market failure. It is optimal to cross-subsidize across firm types, but this is impossible under competition.

We now turn to a three-period economy. As we shall see, the analysis is considerably more complex and the results quite different and more interesting.

III. The Three-Period Economy

III.1 Characterization of competitive equilibrium and planning optimum

In this section we consider a simple three-period economy. As we indicated earlier, it has the same qualitative efficiency properties as economies with more periods or continuous time.

In our three-period economy, there are three possible employment histories. A worker may stay with the same firm for all three periods of his working life. Or he may quit the firm he started with after one period and stay with the next firm for the remaining two periods. Or he may work for a different firm for each of the three periods of his working life. (We exclude by assumption the possibility that a worker will stay with his first firm for two periods and then join another firm for the third period. This is done in order to simplify notation.)[22] With our definition of a firm, there are as many firm types as employment histories. We shall define a type 1 firm to be one that hires workers when they first join the labor force; a type 2 firm, one that hires workers who have quit a type 1 firm after working for it for one period; and a type 3 firm, one that hires workers who have quit successively their type 1 and type 2 jobs. Thus, a type 1 firm hires workers at the beginning of the first period, a type 2 firm hires workers at the beginning of the second, and a type 3 firm hires workers at the beginning of the third. We employ the same notation as in the previous section.

To simplify the analysis, we assume that tastes are such that capital markets are perfect (see fn. 21b). Later we note how the qualitative results are modified when capital markets are, instead, imperfect.

The three-period economy is qualitatively different from the two-period economy. In the latter economy, firms of neither type could affect the contract offered by firms of the other type. The type 2 firm wage was determined by its budget constraint. And type 1 firms could not alter the wage offered by type 2 firms. The employment contracts of the two types of firms were therefore what we term independent. In the three-period economy, however, type 1 and type 2 firm contracts are interdependent. Type 2 firms' choice of wage structure can affect the contract chosen by type 1 firms, and vice versa. It is this interdependence that gives rise to several market failures.

We begin by characterizing the competitive equilibrium. As for the two-period economy, it is derived as the solution to a two-stage maximization problem. In the first stage, the worker makes his optimizing decision, treating the vector of wages as fixed. In the second stage, each firm chooses its contract, taking into account how workers respond to changes in its wage structure, but treating the wages offered by all other firms as fixed.

III.1.1 Worker's problem

In the two-period problem, we expressed the worker's utility as a function of the wages and levels of job satisfaction for the two periods. In so doing, we treated implicitly the individual's decision concerning consumption in the two periods. As we indicated earlier, it will turn out that in the three-period model, the market failure we identify comes about through interdependence of the employment contracts offered by type 1 and type 2 firms via the individual's saving decisions. For this reason, in this section we model the worker's consumption-savings decisions explicitly. We employ the additional notation:

c_i^j consumption in period i if working for a firm of type j

s saving during the first period.

We now need to specify when, during a period, a worker makes his consumption-saving decision. We shall assume that he makes it prior to finding out his job satisfaction in the period, if this is uncertain. The assumption does not affect our qualitative results, and is chosen only to simplify notation.

The worker's maximization is a dynamic programming problem. We therefore work backwards. In the third period of his working life, the worker has no economic decisions to make. He will retire at the end of the period, and his third period consumption equals his third-period wage plus his savings, the magnitude of which is determined from past economic decisions.

Let us now consider the second-period problem of a worker who is employed with a type 2 firm. He has previously decided on first-period savings, s. He must now decide how much to consume during the current period and the minimum level of job satisfaction corresponding to which he will stay with the firm for the third-period. Generally, both these decisions will be affected by the quality of his first-period job match. If he had a very unfortunate experience in the first period, he might as a result be more risk-averse and hence more reluctant to quit his type 2 job than if his experience had been more favorable. Treating this would considerably complicate the notation of the problem without altering any of the efficiency properties of the solution. To circumvent this difficulty, we assume that the worker's direct utility function is strictly separable in all its arguments. To further simplify the analysis, we assume that workers' tastes remain the same over time. Thus,

$$\mathcal{U}\,[\{c_t\},\{\theta_t\}] = \sum_{t=1}^{3} (u[c_t]+v[\theta_t]), \tag{12}$$

where $\mathcal{U}[\cdot]$ is the direct utility function and subscript t denotes the time period. We recardinalize job satisfaction such that $v[\theta_t]\to\theta_t$,[22a] and normalize utility so that the expected job satisfaction from a random job match is zero. We assume additionally that $u'[c_t]>0$, $u''[c_t]<0$, and $u[c_t]$ analytic. To further simplify notation, we assume that a worker's job satisfaction remains constant during his period of tenure with a firm.[23]

Since utility is separable over time, a worker's objective, at the time he starts employment with a type 2 firm, is to maximize the expected value of $\sum_{t=2}^{3} (u[c_t]+\theta_t)$ which we denote as $\mathcal{E}\hat{\mathcal{U}}$ and term prospective utility. He will quit the type 2 firm at the end of period 2 if his job satisfaction is less than $\bar{\theta}_2^2$. The probability that he quits is therefore $F[\bar{\theta}_2^2]$. If he quits, his third-period consumption is $w_2^2+w_3^3+s-c_2^2$, while if he does not, it is $w_2^2+w_3^2+s-c_2^2$. His expected prospective utility from job satisfaction is $\int_{-\infty}^{\infty} \theta_2^2\, dF[\theta_2^2] = 0$ (period 2) plus $\int_{-\infty}^{\infty} \theta_3^3\, dF[\theta_3^3]\, F[\bar{\theta}_2^2] = 0$ (expected period 3 job satisfaction if he quits times the probability of quitting) plus $\int_{\bar{\theta}_2^2}^{\infty} \theta_2^2\, dF[\theta_2^2]$ (expected period 3 job satisfaction if he does not quit times the probability of his not quitting). The maximization problem facing a worker at the time he joins a type 2 firm is therefore

$$\max_{c_2^2, \bar{\theta}_2^2} \hat{\mathcal{E}\mathcal{U}} = u[c_2^2] + F[\bar{\theta}_2^2]u[w_2^2+w_3^3+s-c_2^2]$$
$$+ (1-F[\bar{\theta}_2^2])u[w_2^2+w_3^2+s-c_2^2] + \int_{\bar{\theta}_2^2}^{\infty} \theta_2^2 \, dF[\theta_2^2]. \tag{13}$$

Where ${}_i^j u$ denotes the utility from consumption in period i for a worker with firm type j, and ' denotes a derivative, the first-order conditions are:

$$c_2^2: \; ({}_2^2u') - (q_2^2 \gamma_3^3 u') - (1-q_2^2 \gamma_3^2 u') = 0, \text{ and} \tag{14a}$$

$$\bar{\theta}_2^2: \quad f[\bar{\theta}_2^2]\,(({}_3^3u) - ({}_3^2u) - \bar{\theta}_2^2) = 0 \;, \tag{14b}$$

which have straightforward interpretations. From (13), (14a), and (14b), we obtain

$$c_2^2 = c_2^2[w_2^2, w_3^2, w_3^3, s], \tag{15a}$$

$$\bar{\theta}_2^2 = \bar{\theta}_2^2[w_2^2, w_3^2, w_3^3, s], \text{ and} \tag{15b}$$

$$\hat{\mathcal{E}\mathcal{U}} = \hat{EU}[w_2^2, w_3^2, w_3^3, s]. \tag{15c}$$

By assumption, the worker who is employed for a type 1 firm for the second period stays there for the third period as well. He will equalize consumption in the two periods.

We move on to the first-period problem facing the worker. He chooses first period saving (or consumption) and the minimum level of job satisfaction corresponding to which he will stay with the firm. Where $\tilde{\mathcal{E}\mathcal{U}}$ is expected utility at the beginning of the first period, his maximization problem is

$$\max_{s, \bar{\theta}_1^1} \tilde{\mathcal{E}\mathcal{U}} = u[w_1^1-s] + 2(1-F[\bar{\theta}_1^1])u[(w_2^1+w_3^1+s)/2]$$
$$+ F[\bar{\theta}_1^1]\hat{EU}[w_2^2, w_3^2, w_3^3, s] + 2\int_{\bar{\theta}_1^1}^{\infty} \theta_1^1 \, dF[\theta_1^1], \tag{16}$$

where $\hat{EU}$ is obtained from (15c). The first-order conditions are:

$$s:\ -({}_1^1u') + (1-q_1^1)({}_2^1u') + q_1^1 \frac{\partial \hat{EU}}{\partial s} = 0, \text{ and} \tag{17a}$$

$$\bar{\theta}_1^1:\ f[\bar{\theta}_1^1](-2({}_2^1u) + \hat{EU} - 2\,\bar{\theta}_1^1) = 0. \tag{17b}$$

And the solution is characterized by

$$s = s[w_1^1, w_2^1, w_3^1, w_2^2, w_3^2, w_3^3], \tag{18a}$$

$$\bar{\theta}_1^1 = \bar{\theta}_1^1[w_1^1, w_2^1, w_3^1, w_2^2, w_3^2, w_3^3], \text{ and} \tag{18b}$$

$$\tilde{EU} = \tilde{EU}[w_1^1, w_2^1, w_3^1, w_2^2, w_3^2, w_3^3]. \tag{18c}$$

We next consider the firms' maximization problems, and subsequently the planner's problem.

III.1.2 Firms' problems

A type 3 firm has no choice. It must pay workers the value of their net marginal product,

$$w_3^3 = m - T^3. \tag{19}$$

A type 2 firm chooses w_2^2 and w_3^2 so as to maximize a worker's prospective utility at the beginning of period 2, $\hat{EU}$, given by (15c), subject to its budget constraint,

$$(2-q_2^2)m - T^2 - (1-q_2^2)w_3^2 - w_2^2 = 0, \tag{20}$$

and the quit rate function. From (15b),

$$q_2^2 = F[\bar{\theta}_2^2] = q_2^2[w_2^2, w_3^2, w_3^3, s]. \tag{21}$$

Substituting (21) into (20) allows us to write the true budget constraint in implicit form as $w_2^2 = A[w_3^2, s]$, which with (18a) gives

$$w_2^2 = A[w_3^2, s[w_1^1, w_2^1, w_3^1, w_2^2, w_3^2]], \tag{22}$$

where $w_3^3 = m - T^3$ is suppressed here and subsequently to simplify notation.

We now have a choice of behavioral assumptions. Will a type 2 firm treat the worker's first-period saving as parametric (exogenous savings), or will it realize and take into account the dependence of this saving on w_2^2 and w_3^2 (endogenous savings)? If a worker knows when he makes his first-period saving decision which type 2 firm he will join if he quits at the end of the period, then each type 2 firm will be able to influence the savings decisions of its entering workers and will take this into account in its calculations. However, if workers, when they quit a type 1 firm, are matched randomly to type 2 firms, or alternatively have to apply (costlessly) to a number of different firms before they find a job vacancy, then type 2 firms will treat the worker's first-period savings as exogenous.

Though both savings assumptions are consistent with our model as specified, we consider the assumption of exogenous savings the more appealing.[24] As a result, we shall focus on exogenous savings in the analysis.[25,26]

With exogenous saving, we shall denote a type 2 firm's perceived budget constraint as

$$w_2^2 = \alpha[w_3^2; s], \tag{23}$$

where the ;s indicates that the firm treats s as exogenous. Note that $\dfrac{\partial \alpha}{\partial w_3^2} = \dfrac{\partial A}{\partial w_3^2}$.

A type 1 firm chooses w_1^1, w_2^1, and w_3^1 so as to maximize a worker's expected utility, $\mathcal{E}\tilde{\mathcal{U}}$, characterized by (18c), subject to its budget constraint,

$$(3-2q_1^1)m - T^1 - (1-q_1^1)(w_2^1+w_3^1) - w_1^1 = 0. \tag{24}$$

Since there are perfect capital markets and since, by assumption, workers are unable (or choose not) to quit type 1 firms at the end of the second period, then only the sum of wages $w_2^1 + w_3^1$ matters, not the particular value of w_2^1 and w_3^1. Without loss of generality, therefore, we set $w_2^1 = w_3^1$. When we subsequently refer to (16), (18a) - (18c), and (22), we shall omit w_3^1 and take w_2^1 to denote both the second- and third-period type 1 wage. Substituting (18a) - (18c) into (24) gives the <u>true</u> budget constraint

$$w_1^1 = B[w_2^1, w_2^2, w_3^2]. \tag{25}$$

A type 1 firm takes w_2^2 and w_3^2 as parametric, and consequently perceives its budget constraint as

$$w_1^1 = b[w_2^1; w_2^2, w_3^2]. \tag{26}$$

Note that $\dfrac{\partial b}{\partial w_2^1} = \dfrac{\partial B}{\partial w_2^1}$.

We are now in a position to characterize the competitive equilibrium.

III.1.3 <u>Characterization of competitive equilibrium</u>

A type 2 firm's problem is to maximize (15c) subject to (23), i.e.

$$\max_{w_3^2} \hat{\mathcal{E}\mathcal{U}} = \hat{EU}[\alpha[w_3^2; s], w_3^2; s], \tag{27}$$

which gives

$$\frac{\partial \hat{EU}}{\partial w_2^2}\frac{\partial \alpha}{\partial w_3^2} + \frac{\partial \hat{EU}}{\partial w_3^2} = 0. \tag{28}$$

If (28) were written less compactly, it would be clear that it has an interpretation analogous to (8a) and (8b); w_3^2 is chosen to equalize the (perceived) marginal revenue and (perceived) marginal cost of providing more implicit insurance in the employment contract.

A type 1 firm's problem involves maximizing (18c) subject to (26); i.e.,

$$\max_{w_2^1} \tilde{\mathcal{E}\mathcal{U}} = \tilde{EU}[b[w_2^1;w_2^2,w_3^2],w_2^1;w_2^2,w_3^2], \tag{29}$$

the first-order condition of which is

$$\frac{\partial \tilde{EU}}{\partial w_1^1}\frac{\partial b}{\partial w_2^1} + \frac{\partial \tilde{EU}}{\partial w_2^1} = 0. \tag{30}$$

To ascertain whether or not there is market failure, we shall investigate whether the maximization conditions characterizing the competitive equilibrium are consistent with those characterizing the planning optimum.

We now turn to the planner's problem.

III.1.4 Characterization of planning optimum

The planner faces only one economy-wide resource constraint, while in the competitive economy there are three different zero-profit constraints, one for each firm type. We argued earlier, for the two-period case with perfect capital markets, that the larger number of constraints in the competitive economy "made no difference". We now establish an analogous proposition for the three-period economy under consideration.

With perfect capital markets, the worker is concerned only with the sum of wages, not their timing. In the economy treated in this section, there are three sums of wages, $W^1 \equiv w_1^1+2w_2^1$, $W^2 \equiv w_1^1+w_2^2+w_3^2$, and $W^3 \equiv w_1^1+w_2^2+w_3^3$, where the W's are indexed by the firm type in which the worker completes his working life. We wish to show that the triple $(\bar{W}^1,\bar{W}^2,\bar{W}^3)$ characterizing the planning optimum is attainable in a competitive economy.[26a] Faced with this triple, workers

have all the information they need in deciding whether or not to quit (and on their contingent consumption streams). Thus, a unique pair of quit rates $\bar{q}_1^1$ and $\bar{q}_2^2$ is associated with this triple. Now set $w_3^3 = m-T^3$, $w_3^2 = w_3^3+\bar{W}^2-\bar{W}^3$, $w_2^2 = (2-\bar{q}_2^2)m-(1-\bar{q}_2^2)w_3^2-T^2$, $2w_2^1 = w_2^2+w_3^2+\bar{W}^1-\bar{W}^2$, and $w_1^1 = (3-2q_1^1)m-T^1-2(1-q_1^1)w_2^1$. It is easily checked that the resulting set of wages is consistent with the triple and satisfies the three zero-profit constraints.[27]

The fact that the planning optimum is attainable in the competitive economy has two important implications. The first is that there is no cross-subsidization market failure in our three-period economy when capital markets are perfect. This market failure arises whenever the planning optimum is not competitively attainable, in which case cross-subsidization across firm types expands the competitive solution set. The second implication is that we may treat the planner as choosing the six wage rates and facing the constraints of the competitive economy. This considerably simplifies the comparison of the planning optimum and the competitive equilibrium.

The planner's problem may therefore be characterized as the maximization of expected utility, given by (18c), subject to the true budget constraints for the three firm types (19), (22), and (25). Treating (19) implicitly, we rewrite (22) as

$$w_2^2 = A[w_3^2, s[w_1^1, w_2^1, w_2^2, w_3^2]]. \qquad (22')$$

And, from before,

$$w_1^1 = B[w_2^1, w_2^2, w_3^2]. \qquad (25)$$

After substituting each of these equations into the other, we obtain

$$w_2^2 = \mathcal{A}[w_2^1, w_3^2], \text{ and} \qquad (31)$$

$$w_1^1 = \mathcal{B}[w_2^1, w_3^2]. \qquad (32)$$

(31) is the zero-profit constraint for the type 2 firm, and (32) the analogous constraint for the type 1 firm. Substituting these into (18c) gives

$$\tilde{\mathcal{EU}} = \tilde{EU}[\mathcal{B}[w_2^1, w_3^2], w_2^1, \mathcal{A}[w_2^1, w_3^2], w_3^2]. \qquad (33)$$

The planner maximizes (33) with respect to w_2^1 and w_3^2, yielding the conditions

$$w_2^1\text{:}\quad \frac{\partial \tilde{EU}}{\partial w_1^1}\frac{\partial \mathcal{B}}{\partial w_2^1} + \frac{\partial \tilde{EU}}{\partial w_2^1} + \frac{\partial \tilde{EU}}{\partial w_2^2}\frac{\partial \mathcal{A}}{\partial w_2^1} = 0, \text{ and} \qquad (34)$$

$$w_3^2\text{:}\quad \frac{\partial \tilde{EU}}{\partial w_1^1}\frac{\partial \mathcal{B}}{\partial w_3^2} + \frac{\partial \tilde{EU}}{\partial w_3^2} + \frac{\partial \tilde{EU}}{\partial w_2^2}\frac{\partial \mathcal{A}}{\partial w_3^2} = 0. \qquad (35)$$

To summarize: The competitive equilibrium (with exogenous savings) is characterized by the equations (14a), (14b), (17a), (17b), (20), (24), (28), and (30); and the planning optimum by the first six of the above, and (34) and (35).

III.2 Comparison of the planning optimum and competitive equilibrium

III.2.1 Perfect capital markets

We now compare the c.e. and the p.o. We first prove that the two do not coincide by showing that the sets of equations characterizing each are inconsistent. And to do this, we substitute the first-order conditions for the c.e., (28) and (30), into the corresponding first-order conditions of the planning optimum, (34) and (35), and prove that the resulting two equations are inconsistent with one another for quadratic utility functions, and are therefore generally inconsistent. The proof provides little economic insight and is therefore relegated to Appendix 1.

We next examine the causes of the market failure, and whether too much or too little implicit insurance is provided in competitive employment contracts. The details of the analysis are messy and are presented in Appendix 2. Here, we discuss only the results and their economic rationale.

We first treat type 1 firms. We shall say that such firms provide 'too much' insurance in their employment contracts, and therefore that there is 'excessive' labor turnover from such firms if, at the competitive equilibrium, the 'true' budget constraint, in w_2^1-w_1^1 space (w_1^1 on the vertical axis), is flatter than the 'true' indifference curve (assuming both to be negatively-sloped). This statement needs to be qualified since, in making this comparison, we hold w_2^2 and w_3^2 at their competitive equilibrium levels (hence the ' ' above).

We first investigate how a typical type 1 firm misperceives its indifference curve, and then how it misperceives its budget constraint. We may then ascertain whether it provides too little or too much insurance.

A. A type 1 firm's misperception of workers' indifference curves

To determine the slope of the 'true' indifference curve at the competitive equilibrium, we trace through the effect on expected utility from all type 1 firms simultaneously raising w_1^1 by one unit, holding w_2^1 fixed, and then from raising w_2^1 by one unit, holding w_1^1 constant.

In response to a rise in w_1^1, workers will increase their first-period saving so as to smooth their consumption streams. Type 1 firms realize this, but <u>they fail to take into account that, as a result, type 2 firms alter their contracts</u>. This increase in savings can be shown to increase unambiguously the quit rate from type 2 firms, who, to break even, are forced to offer a less attractive contract. Thus, a unit increase in w_1^1 raises workers' expected utility less than type 1 firms think. It can be shown that a unit increase in w_2^1, meanwhile, causes workers to save less, thereby allowing type 2 firms to offer a more attractive contract, and therefore raises workers' expected utility more than type 1 firms think. Putting these two results together implies that type 1 firms perceive workers' indifference curves, in w_2^1-w_1^1 space, to be flatter than is actually the case; i.e., type 1 firms overestimate the marginal benefit to its workers from insurance. This effect, <u>by itself</u>, causes type 1 firms to set w_1^1 too high and w_2^1 too low, thus providing excessive insurance.

B. A type 1 firm's misperception of its budget constraint

Here we perform a similar exercise.

As a result of a unit increase in w_1^1, w_2^1 fixed, workers increase their first-period savings. <u>Type 1 firms fail to realize that this induces type 2 firms to alter their contracts which will in turn affect the quit rate from type 1 firms and hence the cost to them of the unit increase in w_1^1</u>.

Whether type 1 firms under- or over-estimate the cost of raising w_1^1 therefore depends on how an increase in first-period savings alters q_1^1 <u>via the change in the type 2 contract it induces</u>. This can be decomposed into two effects: first, how an increase in first-period savings alters the type 2 contract, and second, how this alteration in the type 2 contract affects q_1^1.

An increase in first-period savings alters the type 2 contract in two ways. First, as we have seen, it causes type 2 firms to offer a less attractive contract. Second, the "income effect" associated with this will result in workers wanting more or less implicit insurance from type 2 firms according to whether their utility functions exhibit decreasing or increasing absolute risk aversion. The net effect of the increase of s on w_2^2 and w_3^2 can be shown to be ambiguous (see Appendix 2). And the effects on q_1^1 of the changes in w_2^2 and w_3^2 induced by the increase in s can be shown to be ambiguous as well (see Appendix 2, and Lemma 5, Appendix 1). Thus, type 1 firms may either under- or over-estimate the cost of raising w_1^1. It can be shown that the cost of raising w_2^1, w_1^1 fixed, too, may be under- or over-estimated by type 1 firms, and that type 1 firms may perceive their budget constraints, in w_2^1-w_1^1 space, to be either flatter or steeper than they 'truly' are; i.e., type 1 firms may under- or over-estimate the marginal cost of insurance.

C. Too much or too little insurance provided by type 1 firms?

We have seen that type 1 firms perceive workers' indifference curves in w_2^1-w_1^1 space to be flatter than is actually the case, and may perceive their budget constraints to be either flatter or steeper than they truly are. The latter misperception can dominate the former. <u>Thus, one cannot say a priori whether type 1 firms provide too much or too little implicit insurance in their employment contracts, nor therefore whether there is too much or too little labor turnover from these firms under competition</u>. Which way the market failure goes depends, in a very complicated way, on workers' risk-aversion characteristics, the size of turnover costs, and the properties of the quit rate functions.

The market failure is the result of both a forward-directed externality and a backward-directed externality. Type 1 firms ignore that, when they perturb their contracts, the change in the amount workers save affects the profitability of type 2 firms - the forward-directed externality. Type 2 firms, in adjusting their contracts to restore budget balance, ignore that this affects how much a worker saves and the profitability of type 1 firms - the backward-directed externality.

A worker faces two statistically independent risks, job satisfaction with a type 1 firm and job satisfaction with a type 2 firm. Despite the fact that these risks are independent, the probabilities of quitting the two types of firms, when insurance is provided, are interdependent. When a type 1 firm provides more insurance, it increases its own quit probability. But this action also affects the amount a worker saves, which in turn influences the probability of a worker's quitting a type 2 firm. We have termed this type of market failure that of seemingly unrelated events. It occurs because events are insured against, rather than the underlying states of nature. Thus, it is a form of market failure that arises only when moral hazard is present.

We now briefly consider the type 2 firm employment contract. Type 2 firms treat the parameters of the type 1 firms' contracts as fixed. By doing so, they ignore that when they simultaneously perturb their contracts, thereby inducing a change in a worker's first period saving, s, type 1 firms will in turn perturb theirs, which will induce a further change in s that will affect the profitability of type 2 firm contracts. Thus, here too there are forward- and backward-directed externalities. With exogenous saving, which we have assumed, a type 2 firm hires workers from the general pool of first-period quitters. If such a firm, by itself, perturbs its contract, it alters the amount saved by its entering workers by only an infinitesmal amount. Thus, each such firm treats s as fixed. However, s changes when all type 2 firms simultaneously perturb their contracts. This effect can be termed a sideways externality, since each type 2 firm neglects that by altering its contract it affects all workers' savings (by an infinitesmal amount) and hence the profitability of other type 2 firms. The net effect of all these externalities is again a market failure of the seemingly unrelated events variety. Furthermore, it can be shown that, as with the type 1 employment contract, one cannot say whether type 2 firms provide too much or too little implicit insurance, nor whether the quit rate from type 2 firms is too low or too high.

A careful reading of the above argument indicates that all of the externalities we have identified operate through s, a worker's first-period savings. To check this, let us return to the type 1 employment contract and trace through the effect of a unit rise in w_1^1, w_2^1 constant, when s is fixed. From (27), a type 2 firm's maximization problem with exogenous saving is

$$\max_{w_3^2} \hat{\mathcal{E}\mathcal{U}} = \hat{EU}[\alpha[w_3^2;s],w_3^2;s].$$

When s is fixed, the profit-maximizing w_3^2 is unaffected by w_1^1 or w_2^1. Thus, there is no forward-directed externality from type 1 to type 2 firms. And since type 2 firms do not alter their contracts, there is no backward- or sideways-directed externality either. The intuitive rationale for this result is as follows: Each type 2 firm maximizes the prospective utility of its entering workers. These workers' pasts affect its choice of contract only via their level of savings. When s is fixed, therefore, the contract chosen by a type 2 firm is independent of its workers' past contracts.

We now consider a small perturbation in type 2 firm contracts that satisfies their perceived budget constraints. Because these contracts maximize expected prospective utility, by the Envelope Theorem this perturbation has a negligible effect on expected prospective utility. From (16) we see that, if s is fixed, w_2^2 and w_3^2 influence $\bar{\theta}_1^1$ and $\mathcal{E}\tilde{\mathcal{U}}$ only via expected prospective utility. Thus, the perturbation affects neither type 1 firms' maximands nor their budget constraints. The externalities are therefore inoperative. If, however, s had been variable, this perturbation would have had a non-negligible effect on s, which would have induced a non-negligible effect on $\bar{\theta}_1^1$, which would have forced type 1 firms to adjust their contracts to restore budget balance, thereby generating the externalities we have identified.

The essential point is that the only effective interdependence between firms that hire workers earlier in their careers and those that hire workers later in their careers comes via savings. Each firm reasons correctly that its own actions have a negligible effect on a typical worker's savings and hence on other firms. But all firms of one type, taken together, have a non-negligible effect on other firms, which gives rise to uninternalized externalities.

III.2.2 Imperfect capital markets

It is difficult to make general statements about how imperfections in capital markets alter the results obtained with perfect capital markets, since the nature of the imperfections can be so varied. There may be no borrowing, lending, or saving; there may be saving, but no borrowing or lending; and when there is borrowing and lending, the interest rate may vary with the amount borrowed or there may be restrictions on the amount borrowed, etc. Furthermore, as we suggested before, the characteristics of capital markets should be derived rather than assumed. It makes little sense to talk about the efficiency properties of an economy in which it is assumed, say, that there is saving, but no borrowing or lending, when in fact there is nothing in the description of the economy to preclude borrowing and lending.

One can nonetheless say something. First, with imperfect capital markets, the cross-subsidization market failure may occur. Second, in a no-savings economy, the externalities we have treated in this section, all of which operate through savings, would not occur, but the cross-subsidization market failure would arise. Finally, in an economy in which capital markets external to the firm are imperfect, the employment contract may be characterized by implicit or informal borrowing and lending.

III.3 Discussion

The model can be generalized to n periods or to continuous time and our simplifying assumptions of a zero interest rate and of an intertemporally separable utility function with unchanging tastes can be relaxed, and the same qualitative results obtained.

Unfortunately, the policy implications of the market failures we have identified are far from clear. One cannot say *a priori* whether firms provide too little or too much insurance in their contracts, nor therefore a priori whether the government should tax or subsidize labor turnover. In principle, one could compute the optimal set of turnover taxes and subsidies for the economy treated in the paper, but the informational requirements to do this are unrealistically large. Besides, one would expect the results to be substantially altered when the realistic complications arising with heterogeneous workers, notably adverse selection problems, are considered. Thus, we judge the results of the paper to be of more theoretical than practical interest.

We tried in the paper to derive, albeit rather casually, the institutional structure of the economy - who will provide insurance in what form, what the characteristics of capital markets will be, what form of market organization will arise, etc. - from its primitive elements (including the information technology), in contrast to almost all other work treating imperfect economies which instead assumes the institutional structure. We think that treating institutional structure as endogeneous is very important and essential to the development of the welfare economics of imperfect economies. As it now stands, many of the supposed market failures in imperfect economies result from model inconsistencies; either the assumed institutional structure would not arise with the information or transactions cost technologies assumed, or else the planner is assumed to act in a way that is inconsistent with these technologies. Probably the least difficult item on this research agenda is the development of a theory of the characteristics of capital markets.

There are many interesting extensions to the model of this paper. Among these are the following:

1. One could treat an economy like the one treated in the paper except that workers differ in, say, their propensity to quit. In this case, both adverse selection and moral hazard problems would occur, and it would be interesting to investigate whether any sources of market failure arise from the interaction of these two phenomena, over and above the sources of market failure that have been identified for each phenomenon separately.
2. The model could be extended to provide a general equilibrium theory of frictional unemployment. Dissatisfied workers quit and randomly search through advertised job vacancies. Search is costly. Firms have U-shaped cost curves, and the number of job vacancies a particular firm advertises depends on the number of workers who have recently quit the firm, the cost of advertising, and the probability of filling a job vacancy. The model should give rise to a rather interesting set of new sources of market failure. When a worker joins the unemployment pool, he imposes an externality on other workers by reducing the probability that an advertised vacancy is unfilled. Similarly, a firm, in advertising a job vacancy, imposes an externality on other firms by reducing the probability of their filling a given job vacancy. A model similar to this in many ways has been developed by Hosios [1980].
3. Another dimension of the quality of a job match could be introduced - a worker's productivity with a particular firm. It would be interesting to characterize the nature of the employment contract in this case. In such a model with risk-neutral workers, Hashimoto [1978] assumes that all workers are paid the same wage and low-productivity workers are fired. But there is nothing in his model to preclude a more efficient form of contract in which workers are paid according to their productivities, after

these have been ascertained by the firm.[28] When workers are risk-averse, the structure of the employment contract is far more complex, and it appears that in certain circumstances it is optimal to fire workers.[29]

4. An obvious extension is to wed the model of the paper to implicit contracts theory, thereby providing a general equilibrium formulation of the theory. Some steps in this direction are taken in Arnott, Hosios, and Stiglitz [1980] and Holmstrom [1981]. The ultimate object is to develop an explanation of involuntary employment in the context of a general equilibrium model. For this it is necessary to explain not only why workers are laid off, but also why these workers cannot obtain employment at the equilibrium spot-market wage.

5. We argue in the next section that there are many social-insurance-related problems that have an economic structure similar to that of our model, and it would be interesting to reformulate the model to treat them.

IV. Concluding Comments

We have obtained a rather striking result for a particular model. Arrow and others have conjectured that when markets are incomplete, institutions arise to fill the holes left by these absent markets in a manner that is efficient, given the transactions cost technology and/or information available. In our model, because of information asymmetries, complete insurance was not available against the underlying uncertainty, the quality of job matches. We argued that in these circumstances the accommodating institution which would arise is the firm providing implicit insurance against quitting in the employment contract. We demonstrated that in these circumstances each firm's choice of contract affects the profitability of all other firms through worker savings. Because firms ignore this interdependence, there are uninternalized externalities which cause market failure. This result is a counterexample to Arrow's conjecture. If the source of market failure we have identified is general, the disturbing implication which follows is that the wide range of public, charitable, and private quasi-insurance institutions in existence are collectively inefficient.

How general in fact are the sources of market failure we identified? They are generic, of a class we termed seemingly unrelated events, and arise whenever the provision of insurance is characterized by moral hazard and individuals obtain insurance from more than one source. This is true whether the insurance is provided by insurance companies or the government or social institutions - fraternal and charitable organizations, family, friends etc. Each provider of insurance will generally ignore that the insurance or quasi-insurance he provides will affect the probability of accidents against which others provide insurance.

The seemingly unrelated events market failures identified in this paper were of a special type that merit attention in their own right. Suppose, in a dynamic economy, that the probability of an accident at a point in time depends not only on the contemporaneous amount (flow) of insurance provided, but also on the value of some state (or stock) variable which is affected by how much insurance has been provided in the past and how much will be provided in the future. If different agents provide insurance over different periods, market failure will occur. Each agent will ignore the effect the insurance he provides has on the stock variable for periods during which he does not provide the insurance. In the model treated in the paper, this stock variable was savings.[34] It could also be state of health, level of education, or work experience. This type of externality is internalized if a single insurer covers an individual throughout his life. This provides an argument in favor of government intervention for accidents with this structure, since the government, as the largest "firm" in the economy, can provide insurance "from cradle to grave" at lower cost than other firms.

FOOTNOTES

* This paper synthesizes and extends Stiglitz [1975] and Arnott [1980]. Stiglitz would like to thank the U.S. Department of Labor for financial assistance. And Arnott and Stiglitz would like to thank the NSF and the SSHRCC for joint financial support. The paper benefitted from helpful comments by Barry Nalebuff and seminar participants at Chicago, Princeton, and Tel-Aviv. Remaining errors are our responsibility.

1. This phenomenon was discussed by Arrow [1965] in the context of the market for medical care. More recently, it has been discussed in the macroeconomic literature dealing with implicit contracts (Azariadis [1975], e.g.) and in the principal-agent literature (Ross [1973] and Stiglitz [1974], e.g.).

2. A theme which pervaded general equilibrium theory with complete markets was the neutrality of institutions (e.g., the corporate veil), and therefore the indeterminacy and irrelevance of institutional structure. With incomplete markets, however, institutions are not in general neutral, and institutional structure is relevant and determined by economic forces.

3. The seminal paper on the topic is Becker [1962]. In the paper Becker, when discussing the sharing of the costs and benefits of firm-specific training between firm and worker, acknowledges that the employment contract will provide implicit insurance, but elsewhere in his discussion neglects this. Subsequent papers which have extended and formalized Becker's analysis (Parsons [1972], Hashimoto [1978] inter alia) have assumed that workers are risk-neutral.

4. If, as is typically the case, firms can obtain insurance-relevant information more cheaply than can insurance agents external to the firm. We shall return to this point later.

5. Similarly, if capital markets are imperfect, firms can profit by setting up an implicit capital market, paying workers more than their marginal products when their marginal utility of income is high, and less when it is low.

6. This paper was written in an attempt to understand better the nature of the market failures uncovered in the current paper.

6a. We use the term 'first best' to apply to an undistorted economy in which information is symmetric.

7. A worker's level of job satisfaction may be formalized as follows: The worker's tastes concerning his work environment can be characterized by his most preferred point in work environment characteristic space which is, the shell of an n-dimensional sphere. The work environment provided by a particular firm can be characterized by another point in this space. The quality of a job match may be measured by the distance between the two points. And the level of job satisfaction is some monotonic transformation of this distance having the range $(-\infty,\infty)$. If we assume that workers' tastes and firm work environments are both uniformly distributed over the shell of the sphere, then the probability distribution of the level of job satisfaction associated with a random match is independent of the firm and of the worker.

8. In this description of the economy, we allow for the possibility that a worker's job satisfaction may change during his period of tenure with a firm, as determined by some stochastic process.

9. The same qualitative results would obtain if it were assumed instead that the firm knows only the worker's age or years of employment at the time he applies for a job.

10. This requires some qualification. The informational advantage of firms must more than offset any advantage an external agent has from specializing in the provision of insurance.

11. Unless there is one large insurance firm for the entire economy, in which case the externalities we identify are internalized. We rule this out by appeal to administrative diseconomies of scale within a firm, which we discuss shortly.

12. In our moral hazard papers, we state this result as: Exclusivity is a necessary condition for optimality. The result is simply demonstrated.

Consider a very simple economy in which there is a single, risky economic activity. Each identical individual works by himself. If he succeeds, output is one; if he fails, output is zero. His probability of success depends on his effort. The insurer can ascertain whether or not an individual succeeds, but cannot observe the individual's effort or the state of nature. The insurance contract specifies the individual's income if he succeeds and if he fails.

Let e denote effort, π the probability of success, y^i the individual's net income after event i (i=0, failure; i=1, success), and $u^i[y^i,e]$ the individual's event-dependent utility function. From the individual's choice of effort, we obtain $e=e[y^0,y^1]$, which, to simplify, we assume is continuous.

A planner would choose $\{y^i\}$ by maximizing workers' expected utility, subject to the economy-wide resource constraint.

$$\max_{y^0,y^1} (1-\pi[e])u^0[y^0,e]+\pi[e]u^1[y^1,e]$$

$$\text{s.t.} \quad \text{i)} \quad e=e[y^0,y^1]$$

$$\text{ii)} \quad (1-\pi[e])y^0 = \pi[e](1-y^1)$$

An exclusive agent would face the same maximization problem. To see why market failure occurs when the individual can purchase insurance from more than one agent, consider the extreme case where an individual purchases an infinitesimal amount of insurance from an infinite number of agents. Each agent, realizing that the insurance he offers has only an infinitesimal effect, will treat the individual's effort as exogenous. All insurance agents act together like a single agent who ignores the dependence of the individual's effort on y^0 and y^1. Thus, too much insurance is provided.

13. If, for instance, lower first-period job satisfaction resulted in a worker being more risk-averse, a less satisfied worker would choose a second-period sub-contract offering more insurance than that chosen by a more satisfied worker.

The firm's offering a choice of sub-contracts and allowing workers to self-select among them improves efficiency. It does not, however, eliminate the problems arising from the unobservability of job satisfaction. Self-selection imposes the constraints that each worker must prefer the sub-contract he chooses to others being offered, and these constraints prevent the firm from providing full insurance.

14. Our model is consistent with either or both sets of assumptions.

15. These results are contained in Helpman and Laffont [1975], Marshall [1976], Pauly [1974], and Shavell [1979], inter alia. These papers point out that any price-quantity contract may be written equivalently as a premium-payout contract. Alternatively, the insured may be quoted a non-linear price schedule having the property that he will choose the equilibrium price-quantity pair.

16. We use the term wage to refer to remuneration in whatever form. Thus, our "wage" contract may involve paying workers the value of their (net of training costs) marginal product and providing explicit quit insurance.

Note that this characterization of the employment contract rules out the possibility of severance pay, which essentially involves paying a worker different amounts for a period depending on whether or not he quits at the end of the period. This assumption, too, was made to avoid adding cumber to an already (necessarily) cumbersome model. As we proceed, we shall point out the ways in which allowing severance pay would modify our results. Suffice it to remark here that allowing severance pay would not alter our basic inefficiency result.

17. Ordinarily, workers with lower job satisfaction will quit and others stay. However, if a worker quits, his future job satisfaction may be more uncertain than if he stays. It is possible that a reduction in period one job satisfaction makes a worker so much more risk-averse that he is more reluctant to quit. Thus, it may occur that the subset of levels of job satisfaction corresponding to which workers quit is not connected. This is not of consequence in our analysis.

18. There may be several contracts satisfying (8a), (8b), and the corresponding second-order conditions. See Arnott [1980]. To give competitive markets the benefit of the doubt, we assume that the process of convergence to equilibrium is such that the best of these contracts is chosen.

19. It is possible that, with insurance, less senior workers are paid below their marginal product and more senior workers above theirs. This would occur if, for instance, turnover costs were small while job satisfaction and consumption were strongly complementary. In this case, in the absence of insurance it would be the more senior workers who have the higher marginal utility of income, on average, since they have higher average job satisfaction. Since insurance serves to reduce the disparity between the average marginal utility of income of more and less senior workers, its provision will result in a transfer from less to more senior workers.

19a. This statement requires qualification.

If the planner were able to do so, and we have implicitly assumed that he is not, he could do even better by intervening in not only the labor market but also the capital market. (This point is made in Diamond and Mirrlees [1978] in the context of the effect of social security on the retirement decision.) To see this, consider the case of perfect capital markets. If the government were to tax or subsidize savings (raise or lower the interest rate), this would alter the amount a worker saved. This in turn would, for a given set of wages, cause some workers to change their quit decisions and hence would alter the quit rate. Now, we argued in the introduction that as a result of the moral hazard arising from the provision of quit insurance, there would be "excessive" (more than in the first-best optimum) quitting. The government's adjusting the interest rate can reduce the amount of excessive quitting.

Put alternatively: Imagine starting off at the first-best optimum. Suppose: i) that workers suddenly become blind to firms' characteristics, and firms to workers' characteristics and job satisfaction, and ii) that firms continue to pay workers their net marginal products. Then there is a welfare loss relative to the first-best optimum associated with the uninsured job satisfaction risk. Call this loss A. Now let firms provide partial quit insurance. This reduces the size of A, but introduces a new form of welfare loss -- that associated with excessive labor turnover. Call this loss B. What we establish in the paper is that the competitive economy does not minimize A + B. Now let the government introduce differential commodity taxation (in the context of the paper, the goods are consumption in different periods, the relative prices of which can be altered by taxing saving or adjusting the interest rate). Associated with this is yet another form of welfare loss, which we label C. At the social optimum A + B + C is minimized, and this will generally entail some differential commodity taxation.

The result that in an economy with asymmetric information, a welfare improvement can be made by employing differential commodity taxation, is a general one. It is investigated in Arnott and Stiglitz [1981b] in the context of moral hazard.

20. In equilibrium, capital markets will not be perfect unless workers choose not to default. This equilibrium condition ensures, in this context, that workers will not borrow a huge amount the first period, intending not to repay the loan. We return to this point shortly.

21. With perfect capital markets, the competitive type 1 contracts are socially optimal. In this situation, (8a) and (8b) are efficiency conditions with the following interpretation: Consider a worker who is indifferent between quitting and staying. If he quits, his income changes by $m-T^2-w_2^1$. Offsetting this must be a change in expected job satisfaction having a monetary-equivalent value of $w_2^1-m+T^2$. His quitting also involves a turnover cost of T^2. Thus, if this worker quits, there is a social cost of $-(w_2^1-m)$. We have argued that the provision of insurance will typically involve paying less senior workers above their net marginal products, and more senior workers below theirs. Since, in the absence of insurance, $w_2^1=m$, it will normally be the case that $w_2^1<m$. In this case, when insurance is provided, a marginal worker quitting entails a social cost. Relative to the first-best optimum, therefore, there is excessive labor turnover. A unit rise in w_1^1 increases the proportion of quitters by $\frac{\partial q_1^1}{\partial w_1^1}$. Hence, $-\frac{\partial q_1^1}{\partial w_1^1}(w_2^1-m)$ is the increase in the deadweight loss associated with excessive labor turnover from increasing w_1^1 by one unit. The interpretation of (8b) is analogous. (8a) and (8b) together imply that the socially optimal type 1 contract minimizes the sum of the deadweight losses, relative to the first-best optimum, associated with, first, incomplete equalization of workers' marginal utilities of income, and, second, excessive labor turnover.

21a. Allowing severance pay may improve efficiency. It allows the firm to choose a different first period wage for quitters and for stayers; the first period wage for stayers may then be chosen to efficiently smooth their wage stream. Allowing severance pay does not, however, circumvent the cross-subsidization market failure. In an economy with saving, the cross-subsidization market failure would arise only if quitters were liquidity-constrained in the first period -- the quitter would like to transfer some of his income from the second period to the first. Severance pay does not solve this problem. In an economy without saving (so that consumption in a period equals contemporaneous remuneration), the cross-subsidization market failure would arise if m-T were either higher or lower than quitters' desired second-period consumption. In either case, severance pay, which is paid in the first period, would not solve the problem.

21b. Perfect capital markets, according to our definition, obtain if and only if all workers' marginal utilities of consumption in the pay-back period are infinite at zero consumption.

22. It will be evident that the assumption does not affect the qualitative results of the analysis.

In an earlier version of this paper, we did not make this assumption and employed a general way to index firm types: A firm type is characterized by the employment history of its entering workers. An employment history, in turn, may be characterized by a sequence of binary digits. A particular digit applies to a particular period of the worker's life, and whether it takes on the value of one or zero depends on whether the worker was with the same, or with a different, firm in the previous period. Finally, the sequence of binary digits can be converted to a base ten index.

A number of the participants in seminars we gave found this method of indexation difficult. In consequence we have, in this version of the paper, adopted a method of indexation that is more intuitive, but does not generalize.

22a. This recardinalization will, of course, alter the shape of the probability density function for job satisfaction.

23. Earlier, to simplify the analysis, we ruled out the possibility of complex contracts which specify a menu of sub-contracts between which a worker can choose after discovering his current period job satisfaction. It can be shown that in the model of this section, such contracts, even if permitted, would not arise.

24. In justifying the structure of our model, we appealed to the costs to a firm of co-ordinating its plants, and to the costs of a type 1 and type 2 firm co-ordinating their wage setting. It seems reasonable, therefore, to assume that the costs of a type 2 firm negotiating a contingent contract with a type 1 firm worker are prohibitive.

We could modify our model slightly to develop a more persuasive case against endogenous savings. Suppose, realistically, that a worker does not completely know his job-related tastes at the time he starts working, but finds out about them in the course of time. Suppose, furthermore, that some of the job attributes are location-specific. A worker in Crested Butte, Colorado would then have to make a contingent contract with a firm in the South in case he found he suffered from the cold, another contingent contract with a firm in a big city in case he found that he missed cultural amenities, etc. The costs of making such contingent contracts would be very high. Furthermore, if a worker were to make more than one such contingent contract, savings would no longer be completely endogenous. With this modification, the expected job satisfaction of a quitter would be positive rather than zero, but this would not alter our analysis in any essential way.

25. We have worked out the case of endogenous saving and the results are qualitatively the same as with exogenous savings.

26. There is an incentive for a type 2 firm to establish ties with a few type 1 firms, agreeing to employ all those workers who quit these type 1 firms, and also to coordinate the setting of wages, since this completely internalizes the interdependence of these firms. However, such an arrangement is tantamount to integration of type 1 and type 2 firms. We have ruled out such behavior by appealing to the costs of such co-ordination, as in footnote 24.

26a. In the way we have formulated the problem, we have excluded the possibility that workers may choose not to work during some periods. We assume that it is optimal for workers to remain employed throughout their working lives, and also that they will choose to do so in competitive equilibrium.

27. The assumption that it is optimal that workers remain employed through all three periods can be shown to imply that workers, when faced with this set of wages, choose to remain employed, even if some of these wages are negative.

The proposition advanced in this paragraph, that with perfect capital markets the planning optimum is competitively attainable, generalizes to n periods and holds for a general intertemporal utility function.

28. One has to consider whether the firm has an incentive to reveal truthfully to a worker his productivity. This depends on the information available to the worker; if he knows the probability distribution of wages paid by the firm and of worker productivity, he can ascertain whether the firm is being honest or not.

29. Arnott and Stiglitz [1981a] show that in certain circumstances when moral hazard is present, insurance contracts will randomize payouts. In the context of this paper, this would involve randomization of wages. If such randomization is not possible for legal reasons, firing some workers, which indirectly randomizes wages, may be beneficial.

30. We put inverted commas around 'true' since this budget constraint is computed with w_2^2 and w_3^2 at their competitive equilibrium levels.

31. We use this notation to indicate the change in w_3^2 from a change in s, holding w_1^1 and w_2^1 fixed, but allowing w_2^2 to vary.

32. This is not strictly correct. To say that type 1 firms provide too much insurance, one should compare the c.e. and the planning optimum. Here we are just considering the effects of a local change in the neighborhood of the competitive equilibrium.

33. $$\frac{\partial c_2^2}{\partial s} = \begin{vmatrix} q_2^2({}^{3}_{3}u'') + (1-q_2^2)({}^{2}_{3}u'') & f(({}^{2}_{3}u')-({}^{3}_{3}u')) \\ f(({}^{2}_{3}u') - ({}^{3}_{3}u')) & -f \end{vmatrix} \div \Delta_2 .$$

From the expression for Δ_2 given in Lemma 2, it is evident that $\partial c_2^2/\partial s < 1$. And since $\frac{\partial c_3^2}{\partial s} = \frac{\partial c_3^3}{\partial s} = 1 - \frac{\partial c_2^2}{\partial s}$, then $\frac{\partial c_3^2}{\partial s} = \frac{\partial c_3^3}{\partial s} > 0$. $\frac{\partial c_2^2}{\partial s}$ may be positive.

34. Diamond and Mirrlees [1978] resembles this paper in that market failure arises because of moral hazard and through savings in a dynamic economy. However, the type of moral-hazard-related market failure they treat, discussed in fn. 19a, is different from the one we treat. We noted there that if we had assumed the taxation of saving to be feasible in our model, we would have obtained their source of market failure, as well as the one we focused on. Similarly, if the D-M model were enriched to include another source of risk, the insurance against which is characterized by moral hazard, the market failure we focused on would appear. And if their model were further enriched, allowing private pensions to supplement social security, yet another form of market failure would occur -- that discussed in fn. 12.

Appendix 1

Proof that the competitive equilibrium and the planning optimum do not coincide with quadratic utility functions

To begin, we solve for $\frac{\partial \widetilde{EU}}{\partial w_i^j}$ and $\frac{\partial \widehat{EU}}{\partial w_i^j}$. From (16), we obtain

$$\frac{\partial \widetilde{EU}}{\partial w_1^1} = {}^1_1u' + \frac{\partial \widetilde{EU}}{\partial \bar{\theta}_1^1} \frac{\partial \bar{\theta}_1^1}{\partial w_1^1} + \frac{\partial \widetilde{EU}}{\partial s} \frac{\partial s}{\partial w_1^1} .$$

From (17a) and (17b), $\frac{\partial \widetilde{EU}}{\partial \bar{\theta}_1^1} = \frac{\partial \widetilde{EU}}{\partial s} = 0$. Thus, $\frac{\partial \widetilde{EU}}{\partial w_1^1} = {}^1_1u'$. Proceeding similarly, taking into account the results of the worker's maximization problem, we get

$$\frac{\partial \widetilde{EU}}{\partial w_i^j} = ({}^j_iu')\pi_i^j, \tag{A1.1}$$

where π_i^j is the proportion of workers who receive wage w_i^j. Also, from (13), (14a), and (14b),

$$\frac{\partial \widehat{EU}}{\partial w_i^j} = \left(\frac{\pi_i^j}{q_1^1}\right) ({}^j_iu'). \tag{A1.2}$$

Comparing (A1.1) and (A1.2), we see that we may rewrite (28) as

$$\frac{\partial \widetilde{EU}}{\partial w_2^2} \frac{\partial \alpha}{\partial w_3^2} + \frac{\partial \widetilde{EU}}{\partial w_3^2} = 0. \tag{28'}$$

Now subtract (28') (the first-order condition for the c.e. with respect to w_3^2) from (35) (the corresponding first-order condition for the p.o.). This gives

$$\frac{\partial \widetilde{EU}}{\partial w_1^1} \frac{\partial \mathcal{B}}{\partial w_3^2} + \frac{\partial \widetilde{EU}}{\partial w_2^2} \left(\frac{\partial \mathcal{A}}{\partial w_3^2} - \frac{\partial \alpha}{\partial w_3^2}\right) = 0. \tag{A1.3}$$

And subtract (30)(the first-order condition for the c.e. with respect to w_2^1) from (34) (the corresponding first-order condition for the planning problem), which yields

$$\frac{\partial \tilde{EU}}{\partial w_1^1}\left(\frac{\partial \mathcal{B}}{\partial w_2^1} - \frac{\partial b}{\partial w_2^1}\right) + \frac{\partial \tilde{EU}}{\partial w_2^2}\frac{\partial \mathcal{A}}{\partial w_2^1} = 0. \qquad (A1.4)$$

We want to show that (A1.3) and (A1.4) are inconsistent for quadratic utility functions.

We shall subscript partial derivatives on the basis of the order of the arguments in (22') and (25). Thus, $\frac{\partial A}{\partial w_3^2} = A_1$, $\frac{\partial A}{\partial s} = A_2$. $\frac{\partial A}{\partial w_1^1} = A_2 s_1$, $\frac{\partial B}{\partial w_3^2} = B_3$ etc. To obtain $\frac{\partial \mathcal{B}}{\partial w_2^1}$, $\frac{\partial \mathcal{B}}{\partial w_3^2}$, $\frac{\partial \mathcal{A}}{\partial w_2^1}$, and $\frac{\partial \mathcal{A}}{\partial w_3^2}$, we totally differentiate (22') and (25), which yields

$$\frac{\partial \mathcal{B}}{\partial w_2^1} = \frac{(-A_2 s_2 B_2 - B_1 + A_2 s_3 B_1)}{\Delta} \qquad \frac{\partial \mathcal{A}}{\partial w_2^1} = \frac{-(A_2 s_1 B_1 + A_2 s_2)}{\Delta} \qquad (A1.5)$$

$$\frac{\partial \mathcal{B}}{\partial w_3^2} = \frac{-(A_1 B_2 + A_2 B_2 s_4 + B_3 - A_2 s_3 B_3)}{\Delta} \qquad \frac{\partial \mathcal{A}}{\partial w_3^2} = \frac{-(A_2 s_1 B_3 + A_1 + A_2 s_4)}{\Delta},$$

where $\Delta = B_2 A_2 s_1 - 1 + A_2 s_3$.

And we noted earlier that

$$\frac{\partial b}{\partial w_2^1} = B_1 \quad \text{and} \quad \frac{\partial \alpha}{\partial w_3^2} = A_1. \qquad (A1.6)$$

Substituting (A1.5) and (A1.6) into (A1.4) and (A1.3), respectively, gives

$$\gamma_1^1 \frac{\partial \tilde{EU}}{\partial w_1^1} + \gamma_2^2 \frac{\partial \tilde{EU}}{\partial w_2^2} = 0, \text{ and} \qquad (A1.7)$$

$$\delta_1^1 \frac{\partial \widetilde{EU}}{\partial w_1^1} + \delta_2^2 \frac{\partial \widetilde{EU}}{\partial w_2^2} = 0, \tag{A1.8}$$

where

$$\gamma_1^1 = B_2A_2(s_1B_1+s_2) \div \Delta \qquad \gamma_2^2 = A_2(s_1B_1+s_2) \div \Delta$$

$$\delta_1^1 = (A_1B_2+A_2B_2s_4+B_3-A_2s_3B_3) \div \Delta \quad \delta_2^2 = A_2(s_1B_3+s_4+A_1B_2s_1+A_1s_3) \div \Delta.$$

Our assumptions that $f[\theta_i^j]>0$ for $\theta_i^j \varepsilon (-\infty,\infty)$ and that $u'[c_t]>0$ guarantee that $\frac{\partial \widetilde{EU}}{\partial w_i^j} > 0$. Thus, a sufficient condition for (A1.7) and (A1.8) to be inconsistent is that there be a $u[\cdot]$ and an $F[\cdot]$ such that $\gamma_1^1\delta_2^2 \neq \gamma_2^2\delta_1^1$ at the competitive equilibrium. We shall prove that with quadratic utility functions,

$$\gamma_1^1\delta_2^2 - \gamma_2^2\delta_1^1 = A_2(s_1B_1+s_2)(A_1B_2+B_3) \div \Delta > 0.$$

We shall proceed in five steps. First, we shall demonstrate that both type 1 and type 2 firms provide partial insurance. Then we shall show successively that $A_2<0$, $s_1B_1+s_2<0$, and Δ is positive and finite. And finally, we shall prove that $A_1B_2+B_3$ is positive with quadratic utility functions.

<u>Lemma 1</u>: Type 1 and type 2 firms provide partial insurance against quitting.

To prove the lemma, we must demonstrate that neither firm provides either no insurance or full insurance.

Type 2 firm: Using (14a), (20), (23), and (A.1.2), we may rewrite (28) as

$$\left(q_2^2 \left({}_3^3u'\right) + (1-q_2^2)({}_3^2u')\right) \left(-(1-q_2^2) + \frac{\partial q_2^2}{\partial w_3^2} (w_3^2-m)\right)$$

$$+ (1-q_2^2)({}_3^2u') = 0. \tag{28'}$$

If no insurance is provided, then $w_3^2=m$, and for (28') to be satisfied it is necessary that ${}_3^3u'={}_3^2u'$. But with no insurance $w_3^3=m-T^3<w_3^2 \Rightarrow {}_3^3u'>{}_3^2u'$. If, on the other hand, full insurance is provided, then $w_3^3=w_3^2 \Rightarrow {}_3^3u'={}_3^2u'$. For (28') to be satisfied it would then be necessary that $\frac{\partial q_2^2}{\partial w_3^2}(w_3^2-m) = 0$. But since $\frac{\partial q_2^2}{\partial w_3^2} < 0$ (total differentiation of (14a) and (14b), and $f[\cdot]>0$) and $w_3^2=m-T^3$, then $\frac{\partial q_2^2}{\partial w_3^2}(w_3^2-m) > 0$.

The proof for a type 1 firm is analogous.

<u>Corollary</u>: ${}_3^3u' > {}_2^2u' > {}_3^2u'$ and $\hat{EU}_s = {}_2^2u' > {}_1^1u' > {}_2^1u'$ (where $\hat{EU}_s \equiv \frac{\partial \hat{EU}}{\partial s}$ etc.)

These inequalities follow directly from Lemma 1, along with (14a) and (17a).

<u>Lemma 2</u>: $A_2 < 0$ and finite.

From (20),

$$A_2 = \frac{\partial w_2^2}{\partial s} = \frac{f[\bar{\theta}_2^2]\,\frac{\partial \bar{\theta}_2^2}{\partial s}\,(w_3^2-m)}{1-f[\bar{\theta}_2^2]\,\frac{\partial \bar{\theta}_2^2}{\partial w_2^2}\,(w_3^2-m)} \quad . \tag{A1.9}$$

And from (14a) and (14b)

$$\frac{\partial \bar{\theta}_2^2}{\partial s} = \frac{\partial \bar{\theta}_2^2}{\partial w_2^2} = ({}_2^2u')({}_3^2u'-{}_3^3u')f[\bar{\theta}_2^2] \div \Delta_2, \tag{A1.10}$$

$$\text{where } \Delta_2 = \begin{vmatrix} {}_2^2u''+(q_2^2)({}_3^3u'')+(1-q_2^2)({}_3^2u'') & f(({}_3^2u')-({}_3^3u')) \\ f(({}_3^2u')-({}_3^3u')) & -f \end{vmatrix}$$

is finite (from assumptions on u and f) and positive (from the second-order conditions of the worker's maximization problem (13)). From Lemma 1 and its corollaries, $w_3^2<m$ and ${}^2_3u' < {}^3_3u'$. These results, along with (A1.9) and (A1.10) and assumptions on u and f, give $A_2<0$ and finite.

<u>Lemma 3</u>: $s_1B_1 + s_2 < 0$.

From total differentiation of (17a) and (17b), $s_1=-2({}^1_1u'')f \div \Delta_1$, where

$$\Delta_1 = \begin{vmatrix} {}^1_1u'' + \dfrac{(1-q_1^1)({}^1_2u'')}{2} + q_1^1\,\hat{EU}_{ss} & f(\hat{EU}_s-({}^1_2u')) \\ f(\hat{EU}_s-({}^1_2u')) & -2f \end{vmatrix}$$

is positive (from the second-order conditions of the worker's maximization problem (16)). Thus $s_1>0$. From (30) and (A1.1), $B_1=-2(1-q_1^1)({}^1_2u')\div({}^1_1u')<0$. And from (17a) and (17b),

$$s_2 = \begin{vmatrix} -(1-q_1^1)({}^1_2u'') & f(\hat{EU}_s-({}^1_2u')) \\ 2({}^1_2u')f & -2f \end{vmatrix} \div \Delta_1,$$

which since $\hat{EU}_s>{}^1_2u'$ (Corollary to Lemma 1), is negative.

<u>Lemma 4</u>: Δ is finite and positive.

Now, $\Delta=B_2A_2s_1-1+A_2s_3$. The finiteness of Δ is proved by demonstrating that each term is finite, its positiveness from the necessary conditions for a maximum of (33).

<u>Lemma 5</u>: If $u''=-k$ (k a positive constant), then $A_1B_2 + B_3 > 0$.

From (24) (recalling that $w_2^1=w_3^1$)

$$A_1B_2+B_3 = \frac{2(w_2^1-m)f[\bar{\theta}_1^1]\left(\dfrac{\partial\bar{\theta}_1^1}{\partial w_2^2}A_1 + \dfrac{\partial\bar{\theta}_1^1}{\partial w_3^2}\right)}{1-2(w_2^1-m)f[\bar{\theta}_1^1]\dfrac{\partial\bar{\theta}_1^1}{\partial w_1^1}}. \qquad (A1.11)$$

From (28) and (A1.2),

$$A_1 = -(1-q_2^2)({}_3^2u') \div ({}_2^2u'). \qquad \text{(A1.12)}$$

And from (17a) and (17b)

$$\frac{\partial\bar{\theta}_1^1}{\partial w_2^2} A_1 + \frac{\partial\bar{\theta}_1^1}{\partial w_3^2} = \begin{vmatrix} {}_1^1u'' + \dfrac{(1-q_1^1)({}_2^1u'')}{2} + q_1^1\ \hat{EU}_{ss} & -q_1^1(\hat{EU}_{s,w_2^2}A_1 + \hat{EU}_{s,w_3^2}) \\ f(\hat{EU}_s - ({}_2^1u')) & -f(\hat{EU}_{w_2^2}A_1 + \hat{EU}_{w_3^2}) \end{vmatrix} \div \Delta_1$$

(A1.13)

$$\text{and} \quad \frac{\partial\bar{\theta}_1^1}{\partial w_1^1} = -({}_1^1u'')f(\hat{EU}_s - {}_2^1u') \div \Delta_1, \qquad \text{(A1.14)}$$

where Δ_1 is as given in Lemma 3.

From (A1.2) and (A1.12),

$$\hat{EU}_{w_2^2} A_1 + \hat{EU}_{w_3^2} = ({}_2^2u')\left(-\frac{(1-q_2^2)({}_3^2u')}{({}_2^2u')}\right) + (1-q_2^2)({}_3^2u') = 0, \text{ and}$$

$$\hat{EU}_{s,w_2^2} A_1 + \hat{EU}_{s,w_3^2} = -((q_2^2)({}_3^3u'') + (1-q_2^2)({}_3^2u''))\left(\frac{(1-q_2^2)({}_3^2u')}{({}_2^2u')}\right) + (1-q_2^2)({}_3^2u'').$$

Since $u'' = -k$, then

$$\hat{EU}_{s,w_2^2} A_1 + \hat{EU}_{s,w_3^2} = -(1-q_2^2)k\left(-\frac{({}_3^2u')}{({}_2^2u')} + 1\right).$$

Since ${}_2^2u' > {}_3^2u'$ (Corollary to Lemma 1), then

$\hat{EU}_{s,w_2^2} A_1 + \hat{EU}_{s,w_3^2} < 0$. From (A1.13), $\Delta_1 > 0$ and $\hat{EU}_s > {}_2^1u'$, this implies that

$\frac{\partial \bar{\theta}_1^1}{\partial w_2^2} A_1 + \frac{\partial \bar{\theta}_1^1}{\partial w_3^2} < 0$. Also, $\frac{\partial \bar{\theta}_1^1}{\partial w_1^1} > 0$. Substituting these results into (A1.11) and using $w_2^1 - m < 0$ gives $A_1 B_2 + B_3 > 0$.

We have shown that the first-order conditions characterizing the exogenous savings competitive equilibrium and those characterizing the planning optimum are inconsistent for quadratic utility functions. They are therefore generally inconsistent, and thus the exogenous savings c.e. does not generally coincide with the p.o.

Appendix 2

Firms' misperception of their budget constraints and workers' indifference curves

We first investigate the nature of a type 1 firm's misperception of its budget constraint. Second, we examine how it misperceives its workers' indifference curves. Third, we combine these to ascertain whether type 1 firms provide systematically too little or too much insurance in their contracts.

In what follows, we subscript partial derivatives on the basis of (22') and (25). Thus, $A_2 = \frac{\partial w_2^2}{\partial s}$, $s_4 = \frac{\partial s}{\partial w_3^2}$, etc.

A. Type 1 firm's misperception of its budget constraint

Suppose the economy is in competitive equilibrium. To determine the slope, in w_2^1-w_1^1 space, of the 'true' budget constraint,[30] we trace through the effect on a type 1 firm's profits from all type 1 firms simultaneously raising w_1^1 by one unit, holding w_2^1 constant, and then from raising w_2^1 by one unit, holding w_1^1 constant.

The first-round cost per entering worker of raising w_1^1 by one unit is $1 + 2 \frac{\partial q_1^1}{\partial w_1^1} (m-w_2^1)$ (the first term is the direct cost, the second the indirect cost). The unit rise in w_1^1 results in each worker increasing savings by s_1 (s_1>0 (Lemma 3, Appendix 1) since, to smooth his consumption stream, the worker will save part of the increase in first-period income). The type 1 firm realizes this, but since it treats other firms' wages as being fixed, it fails to recognize that this increase in all workers' first-period saving will, by altering q_2^2, result in type 2 firms modifying their contracts. The changes in w_2^2 and w_3^2 will, in turn, affect both q_1^1 and s. Associated with this induced change in q_1^1 is the second-round cost of raising w_1^1. The induced change in s, meanwhile, will result in type 2 firms modifying their contracts yet again, which will generate third-round effects, and so on.

A unit rise in saving causes w_3^2 to change by $\frac{\delta w_3^2}{\delta s}$ (we shall indicate how this is determined shortly)[31] and, in order to maintain budget balance for type 2 firms, w_2^2 to change by $A_2 + A_1 \frac{\delta w_3^2}{\delta s}$ (the first term is the amount by which w_2^2 has to be raised, to restore budget balance, from a unit rise in s, holding w_3^2 fixed; the second term is the additional change in w_2^2 required to restore budget balance as a result of the change in w_3^2 induced by the unit rise in s). Thus, the second-round cost of raising w_1^1 is

$$2s_1 \left(\frac{\partial q_1^1}{\partial w_2^2} (A_1 \frac{\delta w_3^2}{\delta s} + A_2) + \frac{\partial q_1^1}{\partial w_3^2} \frac{\delta w_3^2}{\delta s} \right) (m - w_2^1),$$

while the second-round change in saving is

$$s_1 \left(\frac{\partial s}{\partial w_2^2} (A_1 \frac{\delta w_3^2}{\delta s} + A_2) + \frac{\partial s}{\partial w_3^2} \frac{\delta w_3^2}{\delta s} \right),$$

which induces a third-round change in q_1^1 and saving, and so on. Proceeding by induction and summing gives that the total cost of raising w_1^1 by one unit is $P + \frac{s_1 D}{1-C}$ for $C<1$, where

$$C \equiv \frac{\partial s}{\partial w_2^2} (A_1 \frac{\delta w_3^2}{\delta s} + A_2) + \frac{\partial s}{\partial w_3^2} \frac{\delta w_3^2}{\delta s}, \quad D \equiv 2 \left(\frac{\partial q_1^1}{\partial w_2^2} (A_1 \frac{\delta w_3^2}{\delta s} + A_2) + \frac{\partial q_1^1}{\partial w_3^2} \frac{\delta w_3^2}{\delta s} \right) (m - w_2^1),$$

and $P \equiv 1 + 2(m - w_2^1) \frac{\partial q_1^1}{\partial w_1^1}$.

By an analogous argument, we obtain that the cost of raising w_2^1 by one unit, holding w_1^1 fixed, is $Q + \frac{s_2 D}{1-C}$ for $C<1$, where $Q \equiv (1 - q_1^1) + 2(m - w_2^1) \frac{\partial q_1^1}{\partial w_2^1}$.

Thus, when C<1, the slope of the 'true' budget constraint is

$$\left(\frac{dw_1^1}{dw_2^1}\right)^t_{b.c.} = -\frac{Q + \frac{s_2 D}{1-C}}{P + \frac{s_1 D}{1-C}}, \tag{A2.1}$$

while the slope of the perceived budget constraint is

$$\left(\frac{dw_1^1}{dw_2^1}\right)^p_{b.c.} = B_1 = -\frac{Q}{P}. \tag{A2.2}$$

We know that $s_1>0$, $s_2<0$ (Lemma 3, Appendix 1), Q>0, and P>0. If, as we later assume, the 'true' budget constraint is negatively-sloped and C<1, then from (A2.1) and (A2.2) the 'true' budget constraint is flatter (in w_1^2-w_1^1 space) than the perceived budget constraint if $\frac{D}{1-C} > 0$. This condition has a ready interpretation. $\frac{1}{1-C}$ is the savings multiplier, giving the ratio of the <u>total</u> increase in s (first- plus second- plus third- --- round increases) to the first round increase in s. D gives the single-round increase in q_1^1 induced by a unit increase in s via the change in w_2^2 and w_3^2 it causes. If $\frac{D}{1-C} > 0$, a type 1 firm underestimates the sensitivity of its quit rate to the amount of implicit insurance it provides, and thereby perceives its budget constraint to be steeper than it actually is. We shall discuss the determinants of the signs of D and 1-C later. Before doing so, we perform an analogous exercise to obtain expressions for the slopes of the 'true' and perceived indifference curves at the c.e.

B. Type 1 firm's misperception of workers' indifference curves

The slope of the perceived indifference curve at the c.e. is

$$\left(\frac{dw_1^1}{dw_2^1}\right)^p_{\bar{u}} = B_1 = -\frac{R}{S}, \tag{A2.3}$$

where $R \equiv (1-q_1^1)(_2^1u')$ and $S \equiv {}_1^1u'$. To determine the slope of the 'true' indifference curve at the c.e., we trace through the effect on expected utility from all firms simultaneously raising w_1^1 by one unit, holding w_2^1 constant, and then from raising w_2^1 by one unit, holding w_1^1 constant.

The first-round increase in a worker's expected utility from raising w_1^1 by one unit is ${}_1^1u'$. The unit rise in w_1^1 results in a worker increasing savings by s_1. Though the type 1 firm fails to recognize it, this change in saving induces type 2 firms to alter their contracts. Proceeding as before, we obtain that the second-round increase in expected utility is

$$s_1\left(\frac{\partial \tilde{EU}}{\partial w_2^2}\left(A_1\frac{\delta w_3^2}{\delta s}+A_2\right)+\frac{\partial \tilde{EU}}{\partial w_3^2}\frac{\delta w_3^2}{\delta s}\right),$$

which by (28') equals $s_1 \frac{\partial \tilde{EU}}{\partial w_2^2} A_2$. And the second-round increase in saving is, from before, s_1C. Proceeding by induction and summing, one obtains that for $C < 1$ the total increase in expected utility from raising w_1^1 by one unit is $S + s_1\left(\frac{\partial \tilde{EU}}{\partial w_2^2}\right)\frac{A_2}{1-C}$. Similarly, the total increase in expected utility from raising w_2^1 by one unit is $R + s_2\left(\frac{\partial \tilde{EU}}{\partial w_2^2}\right)\frac{A_2}{1-C}$. Letting $E \equiv A_2 \frac{\partial \tilde{EU}}{\partial w_2^2}$, we get that the slope of the 'true' indifference curve is

$$\left(\frac{dw_1^1}{dw_2^1}\right)_{\bar{u}}^t = -\frac{R+\frac{s_2E}{1-C}}{S+\frac{s_1E}{1-C}} \quad \text{for} \quad C<1. \qquad \text{(A2.4)}$$

R, S, and s_1 are positive, and s_2 is negative. If, as we later assume, the 'true' indifference curve is negatively-sloped and $C < 1$, then from (A2.3) and (A2.4), the 'true' indifference curve is steeper in w_2^1-w_1^1 space, than the perceived indifference curve if $\frac{E}{1-C} < 0$. This, too, has a straightforward interpretation. $\frac{1}{1-C}$ is again the savings multiplier, while E gives the single-round increase in expected utility induced by a unit increase in s via the change in w_2^2 and w_3^2 it causes. If $\frac{E}{1-C} < 0$, a type 1 firm overestimates the sensitivity of expected utility to the amount of implicit insurance it provides, and thereby perceives the indifference curve facing it to be flatter than is actually the case.

C. Type 1 firms - overprovision or underprovision?

To simplify the analysis, we assume in the following discussion that the 'true' budget constraint and the 'true' indifference curve are both negatively-sloped and that C < 1.

We shall say that type 1 firms provide too much insurance in their contracts (relative to the constrained optimum), which results in excessive labor turnover,[3] if at the c.e. the 'true' budget constraint is flatter (less negatively-sloped) than the 'true' indifference curve; i.e.,

$$\left(\frac{dw_1^1}{dw_2^1}\right)_{b.c.}^{t} > \left(\frac{dw_1^1}{dw_2^1}\right)_{\bar{u}}^{t},$$

which from (A2.1) - (A2.4), noting that the perceived slopes are equal, is equivalent to

$$\frac{DS}{1-C} > \frac{EP}{1-C}, \text{ or } \frac{D}{1-C} > \lambda \frac{E}{1-C}, \qquad \text{(A2.5)}$$

where $\lambda = \frac{P}{S}$ is the reciprocal of the marginal utility of period one income.

We now investigate the signs of D, E, and D−λE. Let us first treat D.

Now,
$$D \equiv 2\left(\frac{\partial q_1^1}{\partial w_2^2}\,(A_1\,\frac{\delta w_3^2}{\delta s} + A_2) + \frac{\partial q_1^1}{\partial w_3^2}\,\frac{\delta w_3^2}{\delta s}\right)\,(m-w_2^1)$$

$$= 2f[\bar{\theta}_1^1]\,(m-w_2^1)\left[\left(\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}\,A_1 + \frac{\partial\bar{\theta}_1^1}{\partial w_3^2}\right)\frac{\delta w_3^2}{\delta s} + A_2\,\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}\right]$$

We know that $(m-w_2^1)>0$ (Lemma 1, Appendix 1), $A_2<0$ (Lemma 2, Appendix 1), $\frac{\partial\bar{\theta}_1^1}{\partial w_2^2} > 0$ ((17a) and (17b)), while $\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}A_1 + \frac{\partial\bar{\theta}_1^1}{\partial w_3^2}$ is negative for quadratic utility functions, but generally of uncertain sign (Lemma 5, Appendix 1). $\frac{\delta w_3^2}{\delta s}$ is obtained from total differentiation of (14a), (14b), (20), and (28). The resulting expression involves terms with third partials of the utility function, and is therefore of ambiguous sign. Recall that D is the single-round increase in q_1^1 induced by a unit increase in s via the change in w_2^2 and w_3^2 it causes. The increase in s causes c_2^2 either to rise by less than the rise in s or to fall.[33] In either event, c_3^2 and c_3^3 rise by equal amounts. Since ${}_3^3u' > {}_3^2u'$ (Lemma 1, Appendix 1), then ${}_3^3u$ rises by more than ${}_3^2u$, which from (14b) implies that the quit rate from type 2 firms increases. This causes type 2 firms to run at a deficit. To restore budget balance, they must adjust the wages they offer; specifically, w_3^2 is changed by $\frac{\delta w_3^2}{\delta s}$ and w_2^2 by $A_1\frac{\delta w_3^2}{\delta s} + A_2$. The indeterminacy of the sign of $\frac{\delta w_3^2}{\delta s}$ is not surprising. The increase in s is equivalent to a lump-sum increase in income at the beginning of period 2 for type 2 firm workers. Whether this will result in type 2 firms offering more or less insurance in their contracts depends on how worker risk-aversion changes with income, which explains the presence of third partials in the expression for $\frac{\delta w_3^2}{\delta s}$. Furthermore, whether an increase in the implicit insurance offered in a type 2 contract raises or lowers q_1^1 is *a priori* ambiguous.

Now we turn to $E \equiv A_2 \frac{\partial \hat{EU}}{\partial w_2^2} = A_2 q_1^1 ({}_2^2u')$. Since $A_2<0$ (Lemma 2, Appendix 1), E is unambiguously negative. Recall that E gives the single-round increase in expected utility induced by a unit rise in s via the change in w_2^2 and w_3^2 it causes. We have shown that this unit rise in s initially causes type 2 firms to run at a deficit. To restore budget balance, they have to offer a less favorable contract, which will lower expected utility.

Next we consider

$$C = \frac{\partial s}{\partial w_2^2}\left(A_1 \frac{\delta w_3^2}{\delta s} + A_2\right) + \frac{\partial s}{\partial w_3^2}\frac{\delta w_3^2}{\delta s}$$

$$= \frac{\delta w_3^2}{\delta s}(s_3A_1+s_4) + A_2s_3.$$

Proceeding as in the proof of Lemma 3, Appendix 1, we obtain

$$s_3 = \begin{vmatrix} -q_1^1 \hat{EU}_{s,w_2^2} & f(\hat{EU}_s - {}_2^1u') \\ -f\hat{EU}_{w_2^2} & -2f \end{vmatrix} \div \Delta_1, \text{ which is}$$

of indeterminate sign, while

$$s_3A_1+s_4 = \begin{vmatrix} -q_1^1(\hat{EU}_{s,w_2^2} A_1 + \hat{EU}_{s,w_3^2}) & f(\hat{EU}_s - {}_2^1u') \\ -f(\hat{EU}_{w_2^2} A_1 + \hat{EU}_{w_3^2}) & -2f \end{vmatrix} \div \Delta_1,$$

which is also of indeterminate sign. C is the single-round change in s caused by a unit increase in s via the change in w_2^2 and w_3^2 it induces, and is of ambiguous sign. While C > 1 appears to be a possibility, we have ruled it out by assumption since it is inconsistent with stability of equilibrium.

Finally, we consider the sign of D-λE, which has the same sign as DS-PE. Now $DS = 2({}_1^1u')(m-w_2^1)f[\bar{\theta}_1^1]\left\{\left(\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}A_1 + \frac{\partial\bar{\theta}_1^1}{\partial w_3^2}\right)\frac{\delta w_3^2}{\delta s} + A_2\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}\right\}$, and

$$PE = \left(1 + 2(m-w_2^1)f[\bar{\theta}_1^1]\frac{\partial\bar{\theta}_1^1}{\partial w_1^1}\right)A_2\, q_1^1({}_2^2u').$$ Thus,

$$DS-PE = 2(m-w_2^1)f[\bar{\theta}_1^1]\left\{({}_1^1u')\left(\frac{\partial\bar{\theta}_1^1}{\partial w_2^2}A_1 + \frac{\partial\bar{\theta}_1^1}{\partial w_3^2}\right)\frac{\delta w_3^2}{\delta s} + A_2\left(({}_1^1u')\frac{\partial\bar{\theta}_1^1}{\partial w_2^2} - ({}_2^2u')\frac{\partial\bar{\theta}_1^1}{\partial w_1^1}\, q_1^1\right)\right\} - A_2({}_2^2u')q_1^1,$$

which is of indeterminate sign.

To summarize (with C<1):

i) In w_2^1-w_1^1 space, the 'true' indifference curve is steeper than the indifference curve perceived by type 1 firms.

ii) The 'true' budget constraint is flatter/steeper than the perceived budget constraint if D >/< 0.

iii) i) and ii) imply that type 1 firms provide too much insurance if D>0. In general, however, one cannot say whether type 1 firms provide too little or too much insurance, and therefore whether there is too little or too much labor turnover from these firms. Which way the market failure goes depends, in a very complicated way, on workers' risk-aversion characteristics, the size of turnover costs, and the properties of the quit rate functions.

BIBLIOGRAPHY

Arnott, R.J., "The Structure of Multi-Period Employment Contracts with Incomplete Insurance Markets", Institute for Economic Research, Queen's University, discussion paper #395, 1980.

Arnott, R.J., and J.E. Stiglitz, "Equilibrium in Competitive Insurance Markets: The Welfare Economics of Moral Hazard", mimeo, August, 1981a.

Arnott, R. J., and J. E. Stiglitz, "Moral Hazard and Optimal Commodity Taxation," mimeo, August 1981b.

Arnott, R.J., A. Hosios, and J.E. Stiglitz, "Implicit Contracts, Labor Mobility and Unemployment", mimeo, December 1980.

Arrow, K., Aspects of the Theory of Risk-Bearing (Helsinki: Yrjö Jahnssonin Säätiö, 1965).

Azariadis, C., "Implicit Contracts and Underemployment Equilibria", Journal of Political Economy, 83, December 1975, 1183-1202.

Becker, G., "Investment in Human Capital: A Theoretical Analysis", Journal of Political Economy, 70, Suppl., October 1962, 9-49.

Diamond, P. A., and J. A. Mirrlees, "A Model of Social Insurance with Variable Retirement," Journal of Public Economics, 10, 1978, 295-336.

Hashimoto, M., "Firm-Specific Human Capital as a Shared Investment", Institute for Economic Research, University of Washington, discussion paper 78-13.

Helpman, E., and J.-J. Laffont, "On Moral Hazard in General Equilibrium Theory", Journal of Economic Theory, 10, 1975, 8-23.

Holmstrom, B., "Equilibrium Long-Term Contracts", discussion paper 414, Center for Mathematical Studies in Economics and Management Science, Northwestern University, revision, Jan. 1981.

Hosios, A., "Externalities, Efficiency, and the Implicit Contract Theory of Employment", mimeo.

Marshall, J.M., "Moral Hazard", American Economic Review, 66, 1976, 880-890.

Parsons, D.O., "Specific Human Capital: An Application to Quit Rates and Layoff Rates", Journal of Political Economy, 80, 1972, 1120-1143.

Pauly, M., "Overinsurance and the Public Provision of Insurance: The Roles of Moral Hazard and Adverse Selection", Quarterly Journal of Economics, 88, 1974, 44-62.

Ross, S., "The Economic Theory of Agency: The Principal's Problem", American Economic Review, 63, May 1973, 134-139.

Rothschild, M., and J. Stiglitz, "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information", Quarterly Journal of Economics, 90, November 1976, 629-649.

Shavell, S., "On Moral Hazard and Insurance", Quarterly Journal of Economics, 93, 1979, 541-562.

Spence, M., and R. Zeckhauser, "Insurance, Information, and Individual Action," American Economic Review Papers and Proceedings, 61, 1971, 380-387.

Stiglitz, J., "Incentives and Risk Sharing in Sharecropping", Review of Economic Studies, April 1974, 219-255.

Stiglitz, J., "An Economic Analysis of Labor Turnover", Working Paper No. 53, The Economics Series, Institute for Mathematical Studies in the Social Sciences, Stanford University, February 1975.