Labor Turnover, Wage Structures, and Moral Hazard: The Inefficiency of Competitive Markets

Richard J. Arnott, *Queen's University*

Joseph E. Stiglitz, *Princeton University*

A multiperiod, 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. Because the quality of a job match is unobservable, workers cannot insure against the risk of a bad match. The firm provides implicit insurance against job dissatisfaction, typically by paying workers more than their net marginal products in their early years with the firm and less subsequently. Since the probabilities of the insured-against events (the quit rates over time) are affected by the amount of such insurance provided, this implicit insurance is characterized by moral hazard. Individuals quit when in the absence of insurance they would not. The equilibrium contract balances out efficiency in risk bearing with efficiency in turnover incentives. We show that the equilibrium contract is not (constrained) efficient and indicate why.

This paper is concerned with the relationship between wage structure and labor turnover. We are concerned, in particular, with the question who effectively pays for specific training and hiring costs. Becker (1962) provided the classic solution for this problem when individuals and firms are risk neutral.[1] Individuals pay the costs when they are hired, but the wages (which equal the posttraining value of their marginal product) that they are subsequently paid must be sufficiently high for the individuals to recoup these expenditures (with interest). Under these circumstances, individuals have the correct incentives for moving; when they move, their specific human capital is, in effect, destroyed; hence they should only move when the increase in their productivity, or the increment in the utility they receive from the nonpecuniary aspects of the job, at least compensates for this; when individuals have paid the training costs themselves, and when their current wages fully reflect their total productivity (including the increment in their productivity resulting from their specific training), then they will move only when it is efficient to do so.

On the other hand, when individuals are risk averse, this wage structure imposes a heavy, undesirable risk on workers. It means, in particular, that if, for one reason or another, individuals have to move or find it desirable to move, they must bear a large loss. Assume, for instance, that there are important nonpecuniary aspects of a job that individuals only find out about after paying for their specific training costs. If the individual finds out he is badly matched, efficiency requires that he move; but the entire risk of whether the individual is well or badly matched is borne by the worker. Since workers are more risk averse than firms, this distribution of risk bearing is clearly inefficient.

There is thus a trade-off: the firm can bear some part of the training costs, lowering the wage of the trained worker below his marginal productivity (to pay for the training); but to the extent the firm does this, individuals will not have the correct incentive to move. Workers' productivity exceeds their wages; they move whenever the expected increment in nonpecuniary benefits is enough to offset any wage differences; but this means that some of the time they will actually move to a job for which the increment in nonpecuniary benefits is not enough to compensate for the differences in productivity.

If there were perfect information and it were costless to write complete contracts, the contract would specify the conditions under which individuals could leave or the compensation that they would have to provide the firm if they leave under each specific set of circumstances; such

---

[1] The seminal paper on the topic is Becker (1962). Subsequent papers that have extended and formalized Becker's analysis (e.g., Parsons 1972; Hashimoto 1978) have assumed that workers are risk neutral.

contracts (though they might be interpreted as forms of indentured servitude) would ensure economic efficiency. But perfect information does not exist, and it would be hopeless to attempt to write such complete contracts. The firm thus must design its wage contracts to balance these two considerations: efficiency in labor turnover versus the efficient distribution of risk bearing.

We have just described an example of a wide class of problems that have recently been discussed—called "moral hazard" problems. They arise whenever the provision of insurance affects individuals' incentives (see Pauly 1974; Marshall 1976; Arnott and Stiglitz 1982; Stiglitz 1983). It has long been believed that the private balancing of the trade-off between risk bearing and incentives would lead to an efficient contract: though obviously workers may be worse off and profits may be lower than they would be with first-best insurance, still given the inherent information problems, there may be no government intervention that would effect a Pareto improvement.

This paper has, then, two objectives to characterize within a simple general equilibrium model: the equilibrium contract, and to show that the market equilibrium is not in general Pareto efficient; there do, in general, exist government interventions that would improve welfare.

The intuition behind our results on the inefficiency of the decentralized market equilibrium contracts can be seen as follows. Consider two firms, the second of which hires the workers that quit the first. Both offer long-term contracts with implicit insurance. The second firm, in deciding what 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 that causes the competitive equilibrium to differ from the constrained optimum.

It turns out that this line of reasoning is correct, as far as it goes, but incomplete. Consider first the case where utility is intertemporally separable. There externalities operate only via workers' savings. To see this, perturb the second firm's contract in such a way that it offers slightly more insurance while staying on its budget constraint, and hold workers' first-period savings fixed. A worker will quit the first firm if his expected utility from quitting exceeds his utility from staying. The perturbation does not affect his utility from staying; nor, since the second firm's contract has been chosen to maximize profits, does it affect his expected utility from quitting (i.e., the envelope theorem applies). Thus, when workers' first-period savings are fixed, the perturbation does not affect the first firm's quit rate or, therefore, its profitability. But in fact, the perturbation will affect workers' first-period savings; savings are a form of self-insurance and will be influenced by the amount of insurance

offered by the second-period firm. The change in a worker's first-period savings will, in turn, alter his utility from staying and expected utility from quitting. Thus, the perturbation with savings variable will influence the first firm's quit rate and profitability. Consider next the case where utility is not intertemporally separable and where workers' first-period savings are again fixed, and perform the same perturbation. It will cause workers with the second firm to alter their consumption programs; because the utility function is not intertemporally separable, this will affect workers' propensity to quit the first firm. The argument presented in the previous paragraph also fails to recognize that the types of market failures that arise depend on the characteristics of capital markets.[2] It does, however, correctly identify the central issue—whether, in this context, competitive markets will efficiently handle the trade-off between risk bearing and incentives.

The argument above suggests that the equilibrium set of employment contracts in a 2-period economy may be efficient. The second firm, which hires workers who have quit the first firm, employs these workers for only 1 period. It has no choice but to set the wage equal to net marginal product. The uninternalized externalities between the two firms identified above are therefore inoperative. They do, however, arise in an economy of 3 or more periods.

I. The Model

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; and so on. 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 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

---

[2] Just as firms provide insurance to their workers, 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. Thus wage profiles will depend on the nature of risk and capital markets available elsewhere in the economy.

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 the starts employment with any firm is independent of both the firm and the worker. 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 depend on the information technology. We assume that the firm knows a worker's employment history at the time he joins the firm,[3] 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 to provide as simple as possible a general equilibrium model with multiperiod employment contracts, labor mobility, moral hazard, and incomplete insurance markets. The qualitative results to be derived will clearly generalize to more complex and realistic economies.

1. The firm is unable to observe a worker's job satisfaction, but it 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 his observed behavior. In fact, there is no efficient method of information revelation. The firm can do no better than offer quit insurance; thus, insurance markets are incomplete. Since the provision of quit insurance affects the probability of the insured-against event, moral hazard is present. And with moral hazard, if an insured's total insurance purchases are observable, an insurance contract will specify both premium and payout. We assume that the firm provides insurance implicitly by setting each period's wage above or below the marginal product.[4]

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

---

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

[4] More complex contracts are possible (see Arnott and Stiglitz 1981). Treating them would complicate the analysis and not alter the qualitative results.

join the firm with a particular employment history. Cross-subsidization is not possible. Suppose it were otherwise and a firm could use the profit from the insurance provided to a lucky group (e.g., 50-year-olds who had only one previous job) to subsidize the insurance provided to an unlucky group (e.g., 30-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. This implies an analytically convenient definition of a firm, which we adopt, as a group of workers in some job who had the same employment history when they started the job.[5]

3. 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 that 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.

4. The moral hazard problem facing the firm is as follows: if it offers more insurance, it bears more of its workers' job match risk, but it also experiences a higher quit rate. The former effect decreases the firm's costs, since workers are willing to take a reduction in average pay in exchange for some insurance; the latter effect, however, increases costs, since workers who quit are subsidized. The firm will trade off these two effects so as to maximize profits.

The social optimum entails a similar trade-off. Increasing the amount of insurance offered reduces worker risk, which is beneficial, but causes excessive turnover, which is harmful.

The central question is, Will a competitive economy be at the optimal point on the worker risk/quit-rate trade-off frontier?

II. The Analysis

For reasons that we have suggested, some of the sources of market failure we identify do not arise in a 2-period economy. The results for the 3-period economy do, however, generalize, and this is the case we treat.

In our 3-period economy, we allow a worker three possible employment histories. A worker may stay with the same firm for all 3 periods of his

[5] If we introduced such realistic complications as the imperfect substitutability of workers in production because of differences in their mix of skills, and economies and diseconomies of scale and scope, we would obtain firms that contained determinate numbers of workers of different types. All we require for the sources of market failure we identify to be operative is that there be some interfirm mobility.

working life. Or he may quit the firm he started with after 1 period and stay with the next firm for the remaining 2 periods. Or he may work for a different firm for each of the 3 periods of his working life. We exclude by assumption the possibility that a worker will stay with his first firm for 2 periods and then join another firm for the third period; this simplifies notation and does not affect any qualitative results. 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 1 period; and a type 3 firm, one that hires workers who have quit successively type 1 and type 2 firms. 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.

The following notation is employed:

w_i^j the wage offered by a firm of type j during the ith period of a worker's life;
θ^j the job satisfaction of a worker in a firm of type j;
$\bar{\theta}^j$ the lowest level of job satisfaction for which a worker will stay with a type j firm;
θ_i the job satisfaction of a worker in period i;
$f[\theta^j]$ the density function of θ^j, with $F[\theta^j]$ the cumulative density function (throughout the paper we use square brackets to enclose the arguments of a function);
s_i^j cumulative saving of a worker employed by a type j firm at the end of the ith period of his working life;
$U[\cdot]$ utility function;
T^j the costs to a firm of type j of hiring and training a worker
c_i consumption of a worker in period i; and
m marginal product of labor.

We make the following simplifying assumptions:

i) For each employment history, workers are indifferent between wage streams with the same present discounted value (PDV) of wages. We term this the perfect capital markets case.[6] (Later we shall comment on how our results are modified when capital markets are imperfect.)

<hr />

[6] We may say that capital markets are imperfect when (and only when) workers are concerned about the timing of their income. This is not a primitive definition (we hope to derive one in subsequent work) but in the context of the paper will suffice. According to this definition, capital markets are imperfect if borrowing or lending rates differ, or if the individual is constrained in the amount he can lend or borrow. Imperfect capital markets may arise because of regulations or the nature of the transactions technology (e.g., administrative costs

ii) $f[\theta^j] > 0$ for $\theta^j \in (-\infty, \infty)$ and is continuous.

iii) $U[\cdot] = \Sigma_{i=1}^3(u[c_i] + \theta_i)$, and $u[\cdot]$ is strictly concave and twice differentiable. Utility is additively separable in consumption and job satisfaction, and over time.[7] Furthermore, job satisfaction is measured by its utility.

iv) The expected job satisfaction from a random job match is zero.

v) In his first period with a firm, a worker must make his consumption decision before finding out his job satisfaction.[8]

vi) The interest rate is zero.

We shall proceed first by characterizing the worker's choice problem, then solving for the (constrained) optimum in which the planner has the same information as firms. Finally, we shall solve for the competitive equilibrium and compare it with the planning solution.

A. The Worker's Problem

At the beginning of the first period of his working life, the worker must decide how much to consume. Then, having discovered how much he likes his job, he must decide whether or not to quit at the end of the period and join another firm. If he decides not to quit, his future decisions are simple; because of the concavity of $u[\cdot]$, he will equalize second- and third-period consumption.

If, instead, he decides to quit at the end of the first period, he must decide how much to consume during the second period. Then, when he finds out the quality of his job match with the second firm, he must decide whether to quit at the end of the second period. In either event, his third-period consumption is determined as a residual.

Under our assumptions, the worker will quit at the end of the first period if his job satisfaction is below some critical level $\bar{\theta}^1$. Similarly, a worker who joins a second firm will quit that firm at the end of the second period if θ is below $\bar{\theta}^2$.

As usual, the intertemporal maximization problem may be set up in a number of different ways. It turns out that for our purposes it is most

in borrowing and lending, the absence of financial assets). More interestingly, they may also result from informational asymmetries; for example, capital markets will typically be imperfect when there is a finite probability of default and when this probability can be affected by actions of the borrower that cannot be observed at reasonable cost by the lender; this is, of course, an example of moral hazard. That perfect capital markets are not inconsistent with our model is argued in Arnott and Stiglitz (1981).

[7] These assumptions result in a worker's choice of employment contract being dependent on his employment history, but not directly on his job satisfaction history.

[8] Relaxing this assumption complicates the analysis by making consumption dependent not only on employment history, but also on realized job satisfaction, and does not alter the qualitative results of the paper.

convenient to treat the worker as choosing the amounts to save and the critical levels of job satisfaction below which to quit (i.e., s_1^1, s_2^2, $\bar{\theta}^1$, and $\bar{\theta}^2$) so as to maximize expected utility. In making this choice, the worker takes the parameters of the employment contracts offered as given. If firms of the same types are offering different contracts, he will choose the contract that maximizes his expected utility.

A worker's expected utility is calculated as the probability that he will stay with the same firm for 3 periods times expected utility conditional on this employment history, plus the probability that he will quit his first firm and stay with his second firm for 2 periods times expected utility conditional on this employment history, plus the probability that he will work for different firms in each of the 3 periods times expected utility conditional on this employment history.

The probability that a worker will stay with the same firm for all 3 periods is the probability that first-period job satisfaction exceeds $\bar{\theta}^1$, $1 - F[\bar{\theta}^1]$. Expected utility is $\sum_{i=1}^{3}(u[c_i] + E\theta_i)$. With this employment history, $c_1 = w_1^1 - s_1^1$, $c_2 = c_3 = (w_2^1 + w_3^1 + s_1^1)/2$, and $E\theta_i = (\int_{\bar{\theta}^1}^{\infty} \times \theta^1 dF[\bar{\theta}^1]) \div (1 - F[\bar{\theta}^1])$ for $i = 1, \ldots, 3$. Since it is only the sum of w_2^1 and w_3^1 that matters to the worker, we set $w_2^1 = w_3^1$. Thus, expected utility contingent on this employment history is

$$u[w_1^1 - s_1^1] + 2u\left[\frac{s_1^1 + 2w_2^1}{2}\right] + 3\frac{\int_{\bar{\theta}^1}^{\infty} \theta^1 dF[\theta^1]}{1 - F[\bar{\theta}^1]}\ .$$

Proceeding similarly for the other two employment histories, we obtain

$$EU = u[w_1^1 - s_1^1] + 2(1 - F[\bar{\theta}^1])u\left[\frac{s_1^1 + 2w_2^1}{2}\right] + 2\int_{\bar{\theta}^1}^{\infty} \theta^1 dF[\theta^1]$$

$$+ F[\bar{\theta}^1](u[w_2^2 + s_1^1 - s_2^2] + (1 - F[\bar{\theta}^2])u[w_3^3 + s_2^2] + \int_{\bar{\theta}^2}^{\infty} \theta^2 dF[\theta^2] \quad (1)$$

$$+ F[\bar{\theta}^2]u[w_3^3 + s_2^2]).$$

Letting $_i^j u$ denote the utility of a worker in period i when with a firm of type j and $_i^j u'$ the corresponding marginal utility of consumption, we may write the first-order conditions of the worker's problem as:

$$s_1^1: -(_1^1 u') + (1 - F[\bar{\theta}^1])(_2^1 u') + F[\bar{\theta}^1](_2^2 u') = 0; \quad (2a)$$

$$s_2^2: F[\bar{\theta}^1](-(_2^2 u') + (1 - F[\bar{\theta}^2])(_3^2 u') + F[\bar{\theta}^2](_3^3 u')) = 0; \quad (2b)$$

$$\bar{\theta}^1: f[\bar{\theta}^1]\{-2(({}^1_2u) + \bar{\theta}^1) + ({}^2_2u) + (1 - F[\bar{\theta}^2])({}^3_2u)$$

$$+ \int_{\bar{\theta}^2}^{\infty} \theta^2 dF[\theta^2] + F[\bar{\theta}^2]({}^3_2u)\} = 0; \tag{2c}$$

$$\bar{\theta}^2: F[\bar{\theta}^1]f[\bar{\theta}^2]\{-(({}^2_3u) + \bar{\theta}^2) + ({}^3_3u)\} = 0. \tag{2d}$$

Equations (2a) and (2b) state that the worker will save so as to equalize the expected marginal utility of consumption. Equations (2c) and (2d) state simply that the worker will quit if doing so increases his expected utility. On the basis of (2a)–(2d) and the budget constraints contained in (1), we can express the worker's savings and quit decisions as functions of the wage vector

$$s^1_1 = s^1_1[w^1_1, w^1_2, w^2_2, w^2_3, w^3_3] \tag{3a}$$

$$s^2_2 = s^2_2[w^1_1, w^1_2, w^2_2, w^2_3, w^3_3] \tag{3b}$$

$$\bar{\theta}^1 = \bar{\theta}^1[w^1_1, w^1_2, w^2_2, w^2_3, w^3_3] \tag{3c}$$

$$\bar{\theta}^2 = \bar{\theta}^2[w^1_1, w^1_2, w^2_2, w^2_3, w^3_3]. \tag{3d}$$

B. The Planner's Problem

We assume that the planner has the same information as the collectivity of firms and can do no better than control wages.[9] Since there are perfect capital markets, all that matters to a worker is the sum of wages for each of the three possible wage histories. We could therefore characterize the planner's problem as the choice of these three wage sums to maximize expected utility, subject to an overall resource constraint. But doing so would make it difficult to compare the planning solution with the competitive equilibrium. We have chosen instead to treat the planner as selecting the wage vector $\{w^j_i\}$. Since there are only three wage sums but five wages, we should be able to specify two additional constraints

[9] 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. 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.

to the planning problem without shrinking the solution set. For reasons that will become apparent, we choose the following two constraints:

$$w_2^2 = (2 - F[\bar{\theta}^2])m - T^2 - (1 - F[\bar{\theta}^2])w_3^2, \qquad (4a)$$

and

$$w_3^3 = m - T^3. \qquad (4b)$$

LEMMA 1. Adding constraints (4a) and (4b) to the planner's maximization problem does not shrink the solution set. See Appendix A for the proof.

The economy-wide resource constraint is

$$w_1^1 + T^1 + 2(1 - F[\bar{\theta}^1])w_2^1 + F[\bar{\theta}^1](T^2 + w_2^2$$
$$+ (1 - F[\bar{\theta}^2])w_3^2 + F[\bar{\theta}^2](T^3 + w_3^3)) - 3m = 0. \qquad (5)$$

Subtracting (4a) and (4b) from (5) gives

$$w_1^1 + T^1 + 2(1 - F[\bar{\theta}^1])w_2^1 - (3 - 2F[\bar{\theta}^1])m = 0, \qquad (6)$$

which we write in compact form as

$$A[w_1^1, \bar{\theta}^1, w_2^1; m, T^1] = 0. \qquad (6')$$

To further simplify notation, we write (1) as

$$EU = EU[s_1^1, s_2^2, \bar{\theta}^1, \bar{\theta}^2, w_1^1, w_2^1, w_2^2, w_3^2, w_3^3], \qquad (1')$$

and (4a) as

$$B(w_2^2, \bar{\theta}^2, w_3^2; m, T^2) = 0, \qquad (4a')$$

and we substitute out for w_3^3 using (4b). Then the planner's problem is to maximize (1') with respect to $(w_1^1, w_2^1, w_2^2, w_3^2)$, subject to equations (4a') and (6'). Let λ be the Lagrange multiplier on (4a') and ϕ the multiplier on (6'). Then the Lagrangian of the planner's problem is

$$\mathcal{L} = EU + \lambda B + \phi A. \qquad (7)$$

The corresponding first-order conditions are

$$w_1^1: \frac{dEU}{dw_1^1} + \lambda \frac{dB}{dw_1^1} + \phi \frac{dA}{dw_1^1} = 0, \qquad (8a)$$

$$w_2^1: \frac{dEU}{dw_2^1} + \lambda \frac{dB}{dw_2^1} + \phi \frac{dA}{dw_2^1} = 0, \qquad (8b)$$

$$w_2^2 : \frac{dEU}{dw_2^2} + \lambda \frac{dB}{dw_2^2} + \phi \frac{dA}{dw_2^2} = 0, \qquad (8c)$$

and

$$w_3^2 : \frac{dEU}{dw_3^2} + \lambda \frac{dB}{dw_3^2} + \phi \frac{dA}{dw_3^2} = 0, \qquad (8d)$$

where dx/dy denotes the derivative of x with respect to y, holding other wage rates fixed but treating the dependence of s_1^1, s_2^2, $\bar{\theta}^1$, and $\bar{\theta}^2$ on y. Equations (8a)–(8d) state simply that the optimal w_1^1, w_2^1, w_2^2, w_3^2 is at a point in $(w_1^1, w_2^1, w_2^2, w_3^2)$ space at which an indifference curve is tangent to the resource constraint.[10]

The planning optimum is therefore characterized by (4a′), (4b), (6′), (8a)–(8d), and (2a)–(2d).

The variable λ is the social benefit in utiles from the planner's receiving \$1.00 per worker employed in type 2 firms from outside the economy; ϕ is the corresponding social benefit when the planner receives \$1.00 per worker employed in type 1 firms. It follows that

$$\lambda = F[\bar{\theta}^1]\phi. \qquad (8e)$$

Equivalently, the economy's resource constraint may be written as $A + F[\bar{\theta}^1]B = 0$, a result we shall use below.

C. Characterization of the Competitive Equilibrium

A type 3 firm has no choice. It must pay workers the value of their marginal product, $w_3^3 = m - T^3$.

A type 2 firm chooses w_2^2 and w_3^2 so as to maximize its profits subject to providing workers with the competitively determined utility level. This utility level will adjust until in equilibrium firms make zero profits. Thus, we may view the firm instead as maximizing a worker's utility, subject to its zero profit constraint, which is given by (4a) or (4a′). In performing this maximization, it takes w_1^1 and w_2^1 as fixed since it treats the equilibrium contract offered by type 1 firms as exogenous. We assume that a type 2 firm gets a random group of workers who have quit type 1 firms. As a result, a single type 2 firm by itself can influence neither the first-period savings of its entering workers nor the quit rate of the firms from which they come, and will therefore treat s_1^1 and $\bar{\theta}^1$ as fixed. It will, however, recognize that its choice of w_2^2 and w_3^2 affects its workers' second-period savings and quit decisions.

---

[10] As pointed out in Arnott (1982), there may be multiple local optima to the planner's and firm's maximization problems. Depending on the path of adjustment to equilibrium, a competitive economy could find itself at an inferior optimum. This is another potential source of market failure.

Thus, a type 2 firm's problem is to maximize (1′) with respect to $\{w_2^2, w_3^2\}$ and subject to (4a′), treating s_1^1, $\bar{\theta}^1$, w_1^1, w_2^1 as fixed. Where $\tilde{\lambda}$ is the Lagrangian multiplier on (4a′), the Lagrangian of the type 2 firm's problem is

$$\mathcal{L} = EU + \tilde{\lambda}B, \qquad (9)$$

and the corresponding first-order conditions are

$$w_2^2: \frac{\delta EU}{\delta w_2^2} + \tilde{\lambda}\frac{\delta B}{\delta w_2^2} = 0, \qquad (10a)$$

and

$$w_3^2: \frac{\delta EU}{\delta w_3^2} + \tilde{\lambda}\frac{\delta B}{\delta w_3^2} = 0, \qquad (10b)$$

where the $\delta x/\delta y$ denotes the derivatives of x with respect to y, holding other wage rates and s_1^1 and $\bar{\theta}^1$ fixed, but treating the dependence of s_2^2 and $\bar{\theta}^2$ on y.

A type 1 firm acts as if it chose w_1^1 and w_2^1 to maximize a worker's expected utility, subject to the zero profit constraint, which is given by (6) or (6′). In performing this maximization, it takes w_2^2 and w_3^2 as fixed, since it treats the equilibrium contract offered by type 2 firms as exogenous, and $\bar{\theta}^1$, $\bar{\theta}^2$, s_1^1, and s_2^2 as varying with w_1^1 and w_2^1 according to (2a)–(2d), since it realizes that its contract choice will affect its workers' first- and second-period savings and quit decisions. Where $\tilde{\phi}$ is the Lagrange multiplier on (6′), the Lagrangian of the type 1 firm's problem is

$$\mathcal{L} = EU + \tilde{\phi}A, \qquad (11)$$

and the corresponding first-order conditions are

$$w_1^1: \frac{dEU}{dw_1^1} + \tilde{\phi}\frac{dA}{dw_1^1} = 0, \qquad (12a)$$

and

$$w_2^1: \frac{dEU}{dw_2^1} + \tilde{\phi}\frac{dA}{dw_2^1} = 0. \qquad (12b)$$

The competitive equilibrium is therefore characterized by (4a′), (4b), (6′), (10a), (10b), (12a), (12b), and (2a)–(2d).

D. Comparison of the Competitive Equilibrium and the Planning Optimum

To provide a thorough treatment of the comparison would greatly lengthen the paper and generate little additional insight. For this reason, we concentrate on what we term the *normal case,* for which, at the social optimum,

$$3m - T^1 > w_1^1 + 2w_2^1 > w_1^1 + w_2^2 + w_3^2 > w_1^1 + w_2^2 + w_3^3 \\ > 3m - T^1 - T^2 - T^3, \tag{13}$$

$m > w_1^1 > m - T^1$, and $m > w_2^2 > m - T^2$. In the normal case, workers who stay with the same firm throughout their working lives receive a sum of wages that is higher than if they were to quit once or twice, but lower than their net marginal product. Workers who quit twice during their working lives receive a sum of wages that is lower than if they were to quit only once or not at all, but higher than their net marginal product. Furthermore, both type 1 and type 2 firms provide partial insurance against quitting, in the sense that these firms cover only part of the worker's hiring costs.

The normal case will certainly obtain when the T^i's are large and workers are only slightly risk averse. Because the T^i's are large, $3m - T^1 \gg 3m - T^1 - T^2 \gg 3m - T^1 - T^2 - T^3$ (where $\gg$ indicates much larger than); and because workers are only slightly risk averse, $m \gg w_1^1 \gtrsim m - T^1$ and $m \gg w_2^2 \gtrsim m - T^2$ (where $\gtrsim$ indicates only slightly larger than), which together imply (13).

LEMMA 2. In the normal case at the social optimum, $({_2^1}u') < ({_1^1}u') < ({_2^2}u')$ and $({_3^2}u') < ({_2^2}u') < ({_3^3}u')$.

For the normal case, the sum of wages when a worker quits once, $w_1^1 + w_2^2 + w_3^2$, exceeds the sum of wages when the worker quits twice, $w_1^1 + w_2^2 + w_3^3$. Thus $w_3^2 > w_3^3$. Now $c_3^2 = s_2^2 + w_3^2$ and $c_3^3 = s_2^2 + w_3^3$. Hence $c_3^2 > c_3^3$, which implies $({_3^3}u') > ({_3^2}u')$. From (2b), the worker chooses s_2^2 to equalize the second-period and expected third-period marginal utilities of consumption, implying $({_3^3}u') > ({_2^2}u') > ({_3^2}u')$.

The proof of the first inequality is similar. For the normal case, the sum of wages when a worker never quits, $w_1^1 + 2w_2^1$ exceeds that when he quits once, $w_1^1 + w_2^2 + w_3^2$. Thus, $2w_2^1 > w_2^2 + w_3^2$. Now $2c_2^1 = 2w_2^1 + s_1^1$ and $c_2^2 + c_3^2 = w_2^2 + w_3^2 + s_1^1$. Hence, $2c_2^1 > c_2^2 + c_3^2$. From above, $c_3^2 > c_2^2$. It follows that $c_2^1 > c_2^2$, and therefore $({_2^1}u') < ({_2^2}u')$. From (2a), the worker chooses s_1^1 to equalize the first-period and expected second-period marginal utilities of consumption, implying $({_2^1}u') < ({_1^1}u') < ({_2^2}u')$. We can now establish

PROPOSITION 1. In the normal case, at the social optimum, competitive type 1 firms will (locally) want to offer additional insurance.

Proof. Hold w_2^2 and w_3^2 at their socially optimal levels, and plot in (w_2^1, w_1^1) space: (i) the social optimum, P; (ii) the indifference contour going through the optimum, $\overline{EU}$; (iii) the true resource constraint, R; and (iv) the budget constraint perceived by type 1 firms, A. Since type 1 firms correctly perceive the indifference contour (recall that they correctly determine how workers will adjust their savings and quit decisions in response to changes in w_1^1 and w_2^1), then to prove the proposition, we need to demonstrate (in terms of fig. 1) that immediately above P, A lies above R.

Let superscript p denote evaluation along the resource constraint, and let c denote evaluation along the budget constraint perceived by competitive firms.

Recalling that $A = 0$ is the type 1 firm budget constraint, we have

$$\left(\frac{dw_1^1}{dw_2^1}\right)_P^c = -\left(\frac{\dfrac{dA}{dw_2^1}}{\dfrac{dA}{dw_1^1}}\right)_P . \tag{i}$$

The social calculation, however, takes account of the fact that changes in w_1^1 and w_2^1 affect the costs incurred on type 2 contracts as well. We demonstrated earlier that the planner faces the budget constraint $F[\bar{\theta}^1]B + A = 0$. Thus
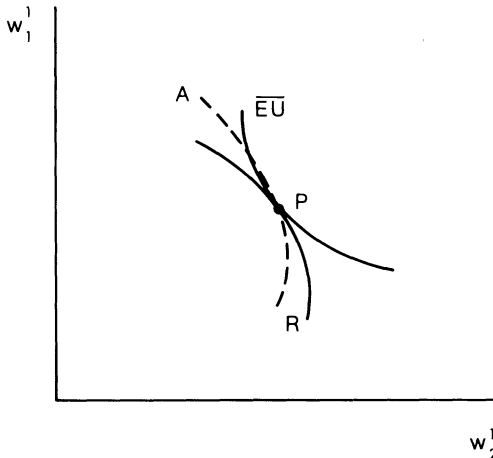

FIG. 1.—The competitive type 1 contract provides "too much" insurance

$$\left(\frac{dw_1^1}{dw_2^1}\right)_P^P = -\left(\frac{F[\bar{\theta}^1]\dfrac{dB}{dw_2^1} + \dfrac{dA}{dw_2^1}}{F[\bar{\theta}^1]\dfrac{dB}{dw_1^1} + \dfrac{dA}{dw_1^1}}\right)_P . \qquad\text{(ii)}$$

It must be the case that $(dEU)/(dw_1^1) > 0$ and $(dEU)/(dw_2^1) > 0$. From (8a) and (8b), it follows that both the numerator and denominator of the term in brackets on the right-hand side of (ii) are negative. And using the results from lemma 2 and those presented in Appendix B, it can be shown that $(dB/dw_1^1) < 0$ and $(dB/dw_2^1) > 0$—raising w_1^1 increases the costs associated with type 2 contracts but raising w_2^1 lowers these costs. (The economics of these and other results will be explained in the next subsection.) Hence, $(dA/dw_1^1)_P \gtreqqless 0$, while $(dA/dw_2^1)_P < 0$. If $(dA/dw_1^1)_P < 0$, then

$$\left(\frac{F[\bar{\theta}^1]\dfrac{dB}{dw_2^1} + \dfrac{dA}{dw_2^1}}{F[\bar{\theta}^1]\dfrac{dB}{dw_1^1} + \dfrac{dA}{dw_1^1}}\right)_P < \left(\frac{\dfrac{dA}{dw_2^1}}{\dfrac{dA}{dw_1^1}}\right)_P , \qquad\text{(iii)}$$

while if $(dA/dw_1^1)_P > 0$, then $(dw_1^1/dw_2^1)_P^c > 0$ and $(dw_1^1/dw_2^1)_P^P < 0$. In either case, at the social optimum, competitive firms would want to increase w_1^1 and decrease w_2^1, thereby providing more insurance. Q.E.D.

We have drawn in the indifference curves and the budget constraints in figure 1 as having standard curvature properties. However, there may be multiple local optima. For this reason, the statement that type 1 firm contracts provide too much insurance is local.

The result in proposition 1 occurs because, for type 1 firms, the private costs of providing insurance are less than the social costs. We shall explain this result in the next subsection. Proposition 1 has an immediate

COROLLARY. In the normal case, the competitive equilibrium set of employment contracts is Pareto inefficient.

PROPOSITION 2. In the normal case, at the social optimum, competitive type 2 firms may want to offer either more or less insurance.

Proof. The proof is analogous to that for proposition 1. Competitive type 2 firms again correctly perceive the indifference contour. Proving the proposition therefore again entails comparing the slopes of the resource constraint and the budget constraint perceived by type 2 firms at the social optimum, but this time in (w_3^2, w_2^2) space. From (4a'), the slope of the budget constraint perceived by type 2 firms is

$$\left(\frac{dw_2^2}{dw_3^2}\right)_P^c = -\left(\frac{\dfrac{\delta B}{\delta w_3^2}}{\dfrac{\delta B}{\delta w_2^2}}\right)_P , \tag{i}$$

while the slope of the resource contraint is

$$\left(\frac{dw_2^2}{dw_3^2}\right)_P^P = -\left(\frac{\lambda \dfrac{dB}{dw_3^2} + \phi \dfrac{dA}{dw_3^2}}{\lambda \dfrac{dB}{dw_2^2} + \phi \dfrac{dA}{dw_2^2}}\right)_P . \tag{ii}$$

After a great deal of tedious manipulation, it can be shown that $-(dw_2^2/dw_3^2)_P^c$ may be either larger or smaller than $-(dw_2^2/dw_3^2)_P^c$. Q.E.D.

These are the basic analytical results. More detail is provided in Arnott and Stiglitz (1981).

E. The Causes of Market Failures: The Normal Case

Let us suppose that w_2^2 and w_3^2 are at their socially optimal values and investigate why a type 1 firm would choose a socially inefficient employment contract (w_1^1, w_2^1).

Consider the firm's calculations with respect to raising w_1^1, holding w_2^1 constant. The firm will take w_2^2 and w_3^2 as fixed. When it raises w_1^1 it causes the worker to alter his savings behavior, changing both s_1^1 and s_2^1. It can be shown that

$$\frac{\partial s_2^2}{\partial w_1^1} = \frac{2(F[\bar{\theta}^1])^2 f[\bar{\theta}^1] f[\bar{\theta}^2](_2^2u'')(_1^1u'')}{\Delta} > 0,$$

where Δ is the determinant of the Hessian corresponding to (2a)–(2d) and is positive. As one would intuitively expect, in response to an increase in w_1^1, the worker will increase his consumption in every period, whatever his employment history, and c_3^2 and c_3^3 are increased by increasing s_2^2. Now, this increase in s_2^2 makes workers more willing to quit type 2 firms. In particular, from (2d), with w_3^2 and w_3^3 fixed, $(d\bar{\theta}^2/ds_2^2) = (_3^3u') - (_3^2u') > 0$ (using lemma 2). In response to their increased quit rates, type 2 firms will be forced to offer a less attractive employment contract in order to restore budget balance. In taking w_2^2 and w_3^2 as fixed, type 1 firms ignore this and hence underestimate the social cost of raising w_1^1.

Now consider the firm's calculations with respect to lowering w_2^1, holding w_1^1 constant. It can be shown that

$$\frac{\partial s_2^2}{\partial w_2^1} = \frac{2F[\bar{\theta}^1]^2 f[\bar{\theta}^2] f[\bar{\theta}^1](\frac{1}{2}u'')\{-(1 - F[\bar{\theta}^1])(\frac{1}{2}u'') + f[\bar{\theta}^1][(\frac{1}{2}u') - (\frac{1}{2}u')](\frac{1}{2}u')\}}{\Delta}$$

$$< 0.$$

Suppose that the worker responds to the fall in w_2^1 by reducing c_2^1 by the full amount of the fall and holding all other $\{c_j^i\}$'s at their original level. This would increase $(\frac{1}{2}u')$, and to restore equality between the first-period and expected second-period marginal utilities of consumption, the worker would have to decrease first-period consumption and increase first-period savings. It can in fact be shown that $(\partial s_1^1/\partial w_2^1) < 0$; thus, in response to the fall in w_2^1, workers joining type 2 firms will have more savings. As a result c_2^2, c_3^2, and c_3^3 will rise, which requires that s_2^2 rise. As we have seen, this increase in s_2^2 forces type 2 firms to offer a less attractive contract. Hence, type 1 firms overestimate the social saving from lowering w_2^1. Putting these two results together, we have that type 1 firms, by ignoring the effects of their contract changes on the profitability of type 2 firms, will set w_1^1 higher and w_2^1 lower than is socially optimal (in a local sense). This entails offering excessive insurance.

We term this uninternalized externality, the *forward-directed externality.* We argued earlier that when the utility function is intertemporally separable, which we have assumed, this externality operates exclusively via workers' savings. To check this, let us retrace the argument of this subsection holding workers' savings fixed. Raise w_1^1, holding w_2^1, w_2^2, and w_3^2 fixed. The increase in w_1^1 alters neither the budget constraints of type 2 firms or the characteristics (savings, tastes) of its entering workers. In consequence, workers will respond in the same way as before to type 2 contracts, and type 2 firms will have no incentive to alter the contracts they offer. The same argument applies to the fall in w_2^1. When the utility function is not intertemporally separable, the past, present, and future are linked not only via savings but also directly, through the utility function.

We now turn to the type 2 firm's calculations. We shall not go through the algebra but instead shall just outline the sources of market failure. There are two of them. First, each type 2 firm, because it receives a random selection of workers, treats s_1^1 as fixed. However, all type 2 firms together alter s_1^1 through the contracts they offer. And s_1^1 in turn affects $\bar{\theta}^2$. Thus, type 2 firms collectively ignore the effect of the contracts they offer on their own quit rates, which operates via first-period savings. We term this the *sideways externality,* because each type 2 firm ignores the effects of the contract it offers on s_1^1 and hence on the profitability of all other type 2 firms. Second, type 2 firms fail to consider that the contract they offer affects the profitability of type 1 firms and hence the attractiveness of the contracts they can offer. This market failure, too,

which we term the *backward-directed externality,* operates via workers' savings.

The forward- and backward-directed externalities are examples of what we have termed elsewhere (Arnott and Stiglitz 1982) the "seemingly unrelated events market failure," which stems from moral hazard. We shall now give an example of this market failure and then show how the problem at hand is analogous to the example. Suppose that there are two statistically independent sets of states of nature. One set influences the probability that one's housing will burn down, the other that of 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's burning down may cause the individual 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. Similarly, the latter firm fails to take into account that the contract it offers will in general affect the profitability of the former firm.

As we suggested earlier, the corresponding 2-period economy is efficient. In such an economy, there are type 1 and type 2 firms. Since type 2 firms hire workers for only 1 period, they have no choice but to pay workers their net (of T) marginal products. Since type 2 firms have no discretion, they cannot affect the profitability of type 1 firms. Nor is their own profitability affected by the type 1 contract.

F. Imperfect Capital Markets

It is difficult to make general statements about how our results are altered when capital markets are imperfect, because there are so many

possible causes of imperfection and so many forms of imperfect capital market. One can, however, identify additional sources of market failure to which imperfections in capital markets may give rise. First, borrowing and lending (whether this is done explicitly by banks or implicitly by firms)[11] results in another market failure of the seemingly unrelated events variety: The borrower or lender will ignore the effect of his actions, through their effects on worker savings, on the profitability of other firms. Second, loans may be characterized by moral hazard; if they are, and if a worker can borrow from different agents at different points in time, or from more than one agent at a point in time, there will be forward-, backward-, and sideways-directed externalities analogous to those discussed in the previous subsection.

Third, a qualitatively different market failure may arise, which we term "the cross-subsidization market failure." Recall that with perfect capital markets, workers were concerned with only the *sum* of wages associated with each of the three employment histories, and that this allowed us to set two wages arbitrarily in the planning problem. The competitive equilibrium had three budget constraints, one for each firm type, while the planning problem had only one resource constraint. But these two extra constraints in the market problem "did not matter," since the planning problem had the two degrees of freedom noted above. With imperfect capital markets, however, the worker is concerned not only with the sum of wages for a particular employment history, but with the timing of wages as well. In this case, the planning problem may have no degrees of freedom and the extra constraints in the market problem may "matter." The potential market failure arises because a planner can subsidize across firm types, but the market cannot; hence the term cross-subsidization market failure. The cross-subsidization market failure is particularly easy to see in the 2-period analogue to the 3-period model treated in the paper. Suppose that expected utility takes on the special form

$$EU = \min(c_1^1, c_2^1, c_2^2) + E\theta_1 + E\theta_2, \tag{i}$$

and that there is no borrowing, lending, or saving. Under competition,

$$c_2^2 = m - T^2 + \eta, \tag{ii}$$

where η is the size of the subsidy to type 2 firms per worker they

[11] This would entail paying workers more in periods when their marginal utility of consumption is high and less in periods when it is low. If younger workers are liquidity constrained, a firm will provide an implicit loan to younger workers that is repaid by older workers who stay with the firm. The older workers who quit the firm in effect default on their implicit loans.

employ, which is collected from type 1 firms. Suppose, furthermore, that in the absence of a subsidy $\min(c_1^1, c_2^1, c_2^2) = c_2^2$. From (i),

$$\left.\frac{dEU}{d\eta}\right|_{\eta=0} = 1,$$

$$\left.\frac{dE\theta_2}{d\eta}\right|_{\eta=0} > 0.$$

(iii)

Thus, in this extreme example, a \$1.00 subsidy payable to type 2 firms for each worker they employ increases *every* worker's utility from consumption by \$1.00.

In sum, while precise characterization will have to await further research, we assert with confidence that with imperfect capital markets, too, multiperiod employment contracts are inefficient.

III. Concluding Comments

A. 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,[12] 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. Thus, the market failures we have identified should be viewed as potential market failures; they warrant government intervention only if the costs of intervention are less than the benefits. Furthermore, one would expect the results to be substantially altered when the realistic complications arising with heterogeneous workers, notably adverse selection problems, are considered.

It is by now well known that in economies with incomplete markets, institutional structure "matters" and is endogenous. One should derive, rather than assume, who will provide insurance in what form, what the characteristics of capital markets will be, what form of market organization will arise, and so on. We have shown in a specific context that the

---

[12] We were able to obtain unambiguous results for type 1 firms. But type 1 firms are special, since they hire workers who have made no prior economic decisions.

institutional structure generated by the market is (potentially) inefficient. Optimal corrective action may entail not only the conventional taxes and subsidies but also the regulation of institutional structure. To treat such considerations, a new welfare economics will have to be developed that is more sophisticated and complex, but at the same time more interesting and realistic, than the old.

What might be a "Chicago economist's" defense of the market against our claim that institutions in a competitive economy may be inefficient? He might point out that the externalities we have identified could be internalized by a superfirm that sets all wages and, in the case of imperfect capital markets, subsidizes across firms, and argue furthermore that the profit motive would lead to the creation of such a superfirm. This is an intellectually respectable defense, since there is nothing in our model to preclude the creation of such a superfirm. It is, however, a rather strange defense of the market that requires that there be one superfirm, which all workers must join for life. One wonders how the forces of competition would operate—potential entry by other superfirms stretches one's credulity—and whether such a superfirm would really be very different from a government. Furthermore, if we were to enrich our model to include many products, and economies and diseconomies of scale and scope (the diseconomies reflecting the costs of coordination and administration), we would find that the optimum entailed a group of superfirms. Competitive equilibrium would be characterized by all the market failures we have identified as long as there is any inter-superfirm mobility of labor. The Chicago economist would probably also stress that we have only identified potential market failures; we have acknowledged this point.

B. Macroeconomic Implications

In this paper we have focused on only one aspect of the contractual arrangements between workers and employers, albeit an important one. The contractual relationship has recently been the center of discussions attempting to explain wage rigidities and unemployment. Elsewhere, we have argued that at least the conventional forms of implicit contract theory do not succeed in providing very persuasive micro foundations for the theory of unemployment (Stiglitz 1984). Though in the presence of perfect information implicit contract theory explains wage rigidities, it does not explain unemployment: workers would still always work up to the point where the value of their marginal productivity was equal to their marginal rate of substitution between goods and leisure. More recent developments, focusing on imperfect information (see, e.g., Grossman and Hart 1981), are equally unconvincing: the contracts typically analyzed entail there being less information than is reasonably available (the terms of the contract are not made contingent on variables, such as

the unemployment rate or imports of cars, etc., which, if the informational problems on which they focus were central, they should be) and more information than is reasonably available (workers must know the total number of hours hired by the firm and, moreover, must monitor transfers of capital goods from the firm to other firms). Under plausible conditions, the contracts lead to overemployment rather than underemployment. Moreover, the contracts are not self-enforcing; to be enforceable, they must be explicit rather than implicit, and, in fact, we see few contracts of the indicated form. The theories do not explain either the patterns of unemployment (i.e., why there are differential unemployment rates among different groups) or forms of unemployment (i.e., why there should be lay-offs rather than work sharing). When there are lay-offs, the theories predict (again under plausible assumptions) that those laid off are better off than those who are retained (provided the firm provides optimal severance pay). Finally, though the theories provide an explanation of layoffs, they do not provide an explanation of unemployment, which must simultaneously explain layoffs and hiring.[13]

The fact that these conventional developments of implicit contract theory have not provided a good theory of unemployment does not, of course, imply that the structure of the contractual arrangements between workers and their employers is *not* an important determinant of unemployment. Several extensions of our model can, we believe, provide at least part of the microfoundations of a theory of unemployment.

It is easy, for instance, to construct a simple model of frictional unemployment, by having a multiperiod model in which there are lags between quitting and being hired and in which the expected duration of unemployment depends on search intensity. Such a model should give rise to a rather interesting set of additional market failures. When a

---

[13] The difference between our implicit contract theory and the standard implicit contract theories with asymmetric information should now be clear; both theories are, in a sense, based on incomplete information. (As we have argued, with perfect information, implicit contracts serve only to smooth demand and do not give rise to any unemployment.) The information problem on which the asymmetric information models have focused concerns the productivity of the firm. Risk-averse firms, wishing to reduce their wages in bad states, need to persuade the workers that the state is in fact bad; to do so, they agree that if they say the state is bad, they will reduce their employment below what it otherwise would be. Because the cost of reducing employment is greater in good states than in bad states, this restriction on employment induces firms to say that it is a bad state only when it is in fact bad. We are concerned with information problems that arise whenever labor turnover is possible. These include problems associated with ascertaining the individual's assessment of the nonpecuniary benefits of the firm (in this paper), and ascertaining the individual's search intensity (in Arnott, Hosios, and Stiglitz 1982).

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, and perhaps also by increasing their quit rates. Furthermore, when a firm chooses its wage structure, it ignores that its quit rate affects the size of the unemployment pool and that a larger unemployment pool may have positive effects on the returns to recruitment by other firms. In addition, the insurance the firm provides against quitting has a moral hazard effect on worker search intensity, which may be harmful to other firms. Whenever a worker moves from one firm to another, the economy's worker-firm matching is altered. Thus, all of these unemployment-related externalities, as well as the externalities noted previously, may be regarded as different forms of matching externalities, which have been discussed elsewhere in the literature.

More central to understanding cyclical unemployment is the extension of the analysis to examine the effect of contractual arrangements on responses to sectoral shocks. Contracts that provide good insurance lead workers to move sluggishly from low- to high-productivity sectors. Firms recognize this in designing contracts; in particular, they recognize that the total cost of the implicit subsidy that they provide in bad states can be reduced if workers can be induced to leave. Workers—ex ante— recognize this and thus are willing to sign contracts that effectively commit them (with some probability) to search in the bad state; the commitment (the *only* enforceable commitment in this context) takes the form of a layoff (possibly with severance pay). Elsewhere Arnott, Hosios, and Stiglitz (1982) have analyzed the structure of these contracts in detail; that analysis provides better micro foundations for the theory of unemployment than existing models: the equilibrium contract may entail layoffs as well as work sharing; the theory is consistent with well-defined patterns of who gets laid off (which are in accord with what is observed); and those who do get laid off are more likely to be worse off than those who are retained (in comparison with the paradoxical results of the standard models).

Though the Arnott-Hosios-Stiglitz model may go some way toward developing a version of implicit contract theory that can provide part of the basis of a theory of unemployment, a still better theory would entail the integration of implicit contract theory with efficiency wage theory. Work on this is underway at present.

C. Conclusion

We have obtained a rather striking result for a particular model. Arrow (1965) and others have conjectured that when markets are

incomplete, institutions arise to fill the holes left by those absent markets in a manner that is efficient, given the transactions cost technology and 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 that would arise is the firm's 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 that cause potential 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 that 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 they 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, and so forth. 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 temporal 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 that 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. 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 through-out his life. To achieve this, either lifetime insurance would have to be provided by a third-party insurer external to firms (which could be the government), or else each individual would have to be employed by the same superfirm throughout his working life that would provide such insurance. The former solution suffers from the problem that information acquisition costs are typically higher for a third-party insurer than for

an individual's employer. The latter solution raises the issue of the compatibility of superfirms and competitive behavior. Both solutions rest on the assumption that individuals are perfectly informed concerning the menu of lifetime contracts offered. How these externalities are best treated when account is taken of incomplete information on the part of individuals is a topic that will have to await future research.

Appendix A
Proof of Lemma 1

We may index employment histories according to the firm type a worker is employed with when he retires. Letting W^k denote the sum of wages associated with employment history k, we obtain

$$W^1 = w_1^1 + 2w_2^1, \tag{A1a}$$

$$W^2 = w_1^1 + w_2^2 + w_3^2, \tag{A1b}$$

and

$$W^3 = w_1^1 + w_2^2 + w_3^3. \tag{A1c}$$

We need to prove that any triple (W^1, W^2, W^3) in the planner's solution set is also in the solution set when (4a) and (4b) are imposed as constraints. First, note that $\bar{\theta}^2$ may be viewed as a function of (W^1, W^2, W^3). Next set

$$w_3^2 = W^2 - W^3 + m - T^3 \tag{A1d}$$

$$w_1^1 = W^3 - (3 - F[\bar{\theta}^2])m + T^2 + T^3 + (1 - F[\bar{\theta}^2])w_3^2, \tag{A1e}$$

and

$$w_2^1 = \frac{(W^1 - w_1^1)}{2} . \tag{A1f}$$

We assume that the utility functions and the exogenous parameters are such that the optimal $\{W_i^j\}$ entails nonnegative $\{w_i^j\}$. It is a straightforward matter to check that (4a), (4b), and (A1d)–(A1f) satisfy (A1a)–(A1c). Furthermore, from (4a), (4b), and (A1d)–(A1f), one can express each wage as a function of (W^1, W^2, W^3) and exogenous parameters. Thus, for any (W^1, W^2, W^3), there is a unique wage vector $(w_1^1, w_2^1, w_2^2, w_3^2, w_3^3)$ satisfying (4a), (4b), and (A1a)–(A1c). Q.E.D.

Appendix B
Evaluation of Derivatives

Total differentiation of (2a)–(2d) gives

$$
\begin{bmatrix}
(_1u'') + (1 - F[\bar\theta^1])(_1u'') + F[\theta^1](_2u'') \\
-F[\bar\theta^1](_2u'') \\
f[\theta^1]((_3u') - (_2u')) \\
0
\end{bmatrix}
$$

$$
=
\begin{bmatrix}
(_1u'') \\ 0 \\ 0 \\ 0
\end{bmatrix} dw_1^1
+
\begin{bmatrix}
-(1 - F[\bar\theta^1])(_1u'') \\ 0 \\ 2f[\theta^1](_1u') \\ 0
\end{bmatrix} dw_2^1
+
\begin{bmatrix}
-F[\theta^1](_2u'') \\ F[\bar\theta^1](_2u'') \\ -f[\theta^1](_2u') \\ 0
\end{bmatrix} dw_1^2
$$

$$
+
\begin{bmatrix}
f[\theta^1]((_2u') - (_2u')) \\
F[\bar\theta^1]((_2u') + (1 - F[\bar\theta^2])(_3u') + F[\theta^2](_3u')) \\
0 \\
F[\theta^1]f[\theta^2]((_3u') - (_3u'))
\end{bmatrix} dw_2^2
$$

$$
+
\begin{bmatrix}
0 \\
F[\theta^1]f[\theta^2]((_3u') - (_2u')) \\
0 \\
-2f[\theta^1]
\end{bmatrix}
\begin{bmatrix}
0 & 0 \\
 & F[\theta^1]f[\theta^2]((_3u') - (_2u')) \\
 & 0 \\
-2f[\theta^1] & -F[\theta^1]f[\theta^2]
\end{bmatrix}
\begin{bmatrix}
ds_1^1 \\ ds_2^2 \\ d\bar\theta^1 \\ d\theta^2
\end{bmatrix},
\quad \text{(B1)}
$$

$$
+
\begin{bmatrix}
0 \\
-f[\theta^1](1 - F[\theta^2])(_3u'') - (_3u') \\
-f[\theta^1](1 - F[\theta^2])(_3u') \\
F[\theta^1]f[\theta^2](_3u')
\end{bmatrix} dw_3^2,
\quad \text{(B2)}
$$

and also

$$\frac{dB}{dw_1^1} = \frac{\partial B}{\partial \bar\theta^2}\,\frac{\partial \bar\theta^2}{\partial w_1^1} = -f[\bar\theta^2](m - w_3^2)\,\frac{\partial \bar\theta^2}{\partial w_1^1}$$

$$\frac{dB}{dw_2^1} = \frac{\partial B}{\partial \bar\theta^2}\,\frac{\partial \bar\theta^2}{\partial w_2^1} = -f[\bar\theta^2](m - w_3^2)\,\frac{\partial \bar\theta^2}{\partial w_2^1}$$

$$\frac{dB}{dw_2^2} = \frac{\partial B}{\partial \bar\theta^2} + \frac{\partial B}{\partial \bar\theta^2}\,\frac{\partial \bar\theta^2}{\partial w_2^2} = -1 - f[\bar\theta^2](m - w_3^2)\,\frac{\partial \bar\theta^2}{\partial w_2^2}$$

$$\frac{dB}{dw_3^2} = \frac{\partial B}{\partial w_3^2} + \frac{\partial B}{\partial \bar\theta^2}\,\frac{\partial \bar\theta^2}{\partial w_3^2} = -(1 - F[\bar\theta^2]) - f[\bar\theta^2](m - w_3^2)\,\frac{\partial \bar\theta^2}{\partial w_3^2}$$

$$\frac{dA}{dw_1^1} = \frac{\partial A}{\partial \bar\theta^1} + \frac{\partial A}{\partial \bar\theta^1}\,\frac{\partial \bar\theta^1}{\partial w_1^1} = -1 - 2f[\bar\theta^1](m - w_2^1)\,\frac{\partial \bar\theta^1}{\partial w_1^1}$$

$$\frac{dA}{dw_2^1} = \frac{\partial A}{\partial w_2^1} + \frac{\partial A}{\partial \bar\theta^1}\,\frac{\partial \bar\theta^1}{\partial w_2^1} = -2(1 - F[\bar\theta^1]) - 2f[\bar\theta^1](m - w_2^1)\,\frac{\partial \bar\theta^1}{\partial w_2^1}$$

$$\frac{dA}{dw_2^2} = \frac{\partial A}{\partial \bar{\theta}^1}\frac{\partial \bar{\theta}^1}{\partial w_2^2} = -2f[\bar{\theta}^1](m - w_2^1)\frac{\partial \bar{\theta}^1}{\partial w_2^2}$$

$$\frac{dA}{dw_3^2} = \frac{\partial A}{\partial \bar{\theta}^1}\frac{\partial \bar{\theta}^1}{\partial w_3^2} = -2f[\bar{\theta}^1](m - w_2^1)\frac{\partial \bar{\theta}^1}{\partial w_3^2}.$$

Using $\partial EU/\partial s_1^1 = \partial EU/\partial s_2^2 = \partial EU/\partial \bar{\theta}^1 = \partial EU/\partial \bar{\theta}^2 = 0$, one obtains

$$\frac{dEU}{dw_1^1} = \frac{\delta EU}{\delta w_1^1} = ({}_1^1 u')$$

$$\frac{dEU}{dw_2^1} = \frac{\delta EU}{\delta w_2^1} = 2(1 - F[\bar{\theta}^1])({}_2^1 u')$$

$$\frac{dEU}{dw_2^2} = \frac{\delta EU}{\delta w_2^2} = F[\bar{\theta}^1]({}_2^2 u')$$

$$\frac{dEU}{dw_3^2} = \frac{\delta EU}{\delta w_3^2} = (1 - F[\bar{\theta}^2])F[\bar{\theta}^1]({}_3^2 u').$$

(B3)

Also

$$\frac{\delta B}{\delta w_2^2} = \frac{dB}{dw_2^2} - f[\bar{\theta}^2](m - w_3^2)\left(\frac{\partial \bar{\theta}^2}{\partial w_2^2} - \frac{\delta \bar{\theta}^2}{\delta w_2^2}\right)$$

$$\frac{\delta B}{\delta w_3^2} = \frac{dB}{dw_3^2} - f[\bar{\theta}^2](m - w_3^2)\left(\frac{\partial \bar{\theta}^2}{\partial w_3^2} - \frac{\delta \bar{\theta}^2}{\delta w_3^2}\right),$$

(B4)

where $(\delta\bar{\theta}^2/\delta w_2^2)$ denotes the change in $\bar{\theta}^2$ from a unit change in w_2^2, all other wages and s_1^1 and $\bar{\theta}^1$ being held fixed, and is calculated using (2b) and (2d).

From the equations above, expressions for all derivatives of interest may be obtained.

461

References

Arnott, R. J. "The Structure of Multi-Period Employment Contracts with Incomplete Insurance Markets." *Canadian Journal of Economics* 15 (1982): 51–76.

Arnott, R. J.; Hosios, A.; and Stiglitz, J. E. "Implicit Contracts, Labor Mobility and Unemployment." Mimeographed. Princeton, N.J.: Princeton University, 1982.

Arnott, R. J., and Stiglitz, J. E. "Labor Turnover, Wage Structures and Moral Hazard: The Inefficiency of Competitive Markets." Econometric Research Program Research Memorandum 289. Princeton, N.J.: Princeton University, September 1981.

———. "Equilibrium in Competitive Insurance Markets: The Welfare Economics of Moral Hazard." Mimeographed. Princeton, N.J.: Princeton University, 1982.

Arrow, K. *Aspects of the Theory of Risk-Bearing*. Helsinki: Yrjo Jahnssönin Säätio, 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 Mirrlees, J. A. "A Model of Social Insurance with Variable Retirement." *Journal of Public Economics* 10 (1978): 295–336.

Grossman, S., and Hart, O. "Implicit Contracts, Moral Hazard, and Unemployment." *American Economic Review Papers and Proceeding* 71 (1981): 301–7.

Hashimoto, M. "Firm-specific Human Capital as a Shared Investment." Discussion Paper no. 78-13. Seattle: University of Washington, Institute for Economic Research, 1978.

Marshall, J. M. "Moral Hazard." *American Economic Review* 66 (1976): 880–90.

Parsons, D. O. "Specific Human Capital: An Application to Quit Rates and Layoffs Rates." *Journal of Political Economy* 80 (1972): 1120–43.

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

Rothschild, M., and Stiglitz, J. E. "Equilibrium in Competitive Insurance Markets: An Essay on the Economics of Imperfect Information." *Quarterly Journal of Economics* 90, no. 4 (November 1976): 629–49.

Stiglitz, J. E. "An Economic Analysis of Labor Turnover." Working Paper no. 53. Stanford, Calif.: Stanford University, Institute for Mathematical Studies in the Social Sciences, February 1975.

Stiglitz, J. E. "Risk, Incentives and Insurance: The Pure Theory of Moral Hazard." *Geneva Papers* 8, no. 26 (January 1983): 4–32. Fifth Annual Geneva Lecture delivered at Zurich, May 1981.

Stiglitz, J. E. "Theories of Wage Rigidity." NBER Working Paper no. 1442. Cambridge, Mass.: NBER, September 1984.