DOCUMENT RESUME

ED 213 808 UD 022 162

AUTHOR Nalebuff, Barry; Zeckhauser, Richard
TITLE Involuntary Unemployment Reconsidered: Second-Best Contracting with Heterogeneous Firms and Workers.
INSTITUTION Wisconsin Univ., Madison. Inst. for Research on Poverty.
SPONS AGENCY Department of Health and Human Services, Washington, D.C.
REPORT NO IRP-DP-675-81
PUB DATE Oct 81
NOTE 68p.; For a related document, see ED 128 596.

EDRS PRICE MF01/PC03 Plus Postage.
DESCRIPTORS Business Cycles; *Economic Climate; Economic Factors; *Employment Practices; Institutional Characteristics; Job Layoff; *Labor Market; Labor Standards; Models; *Productivity; Quality Control; *Unemployment; Wages
IDENTIFIERS *Implicit Contract Theory

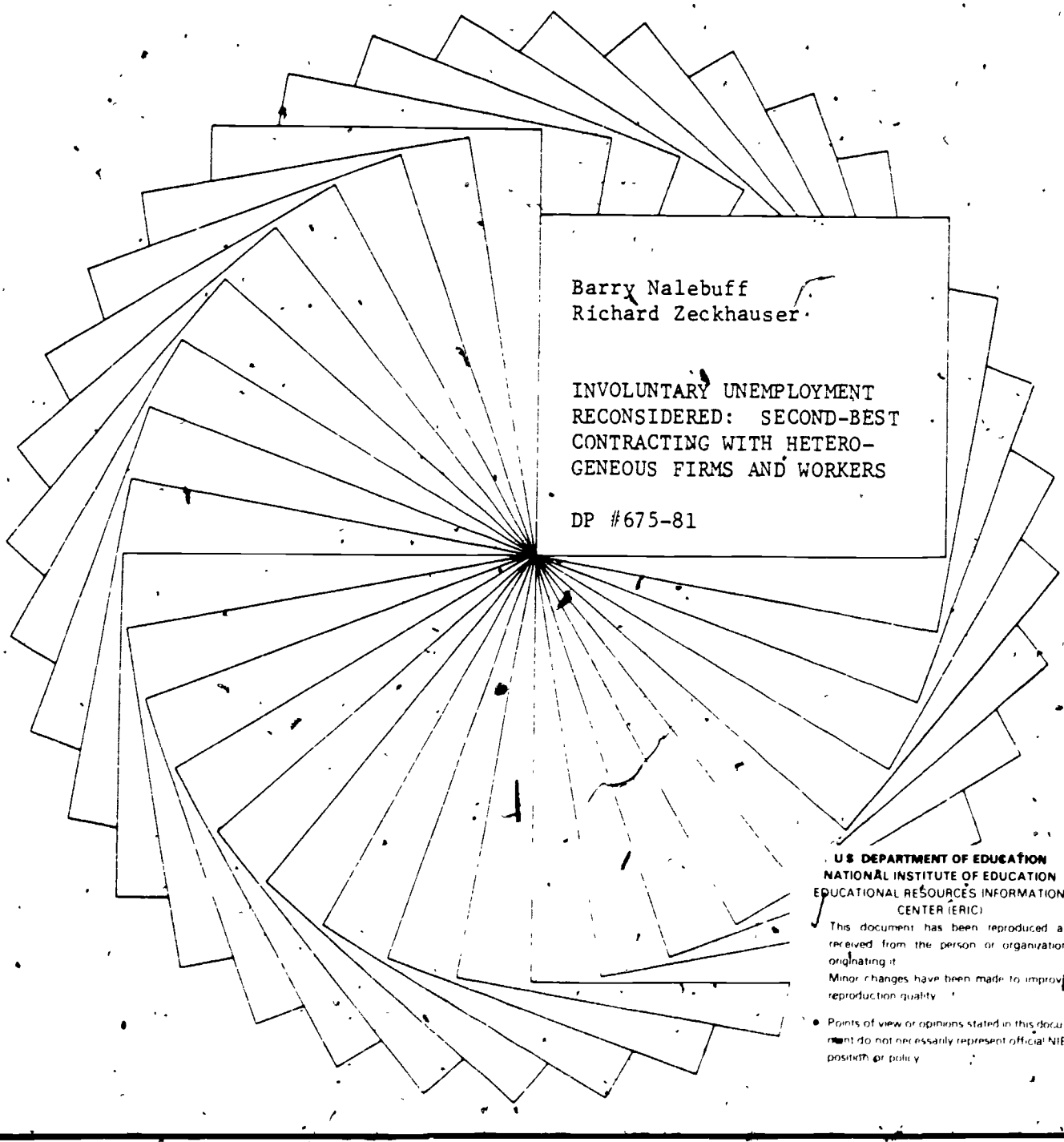
ABSTRACT

The implicit contract theory, a new explanation for the phenomena of involuntary unemployment, does not capture the salient characteristics of real work employment. By building on implicit contract theory, this paper takes into account circumstances ignored in the traditional model: (1) institutional characteristics of the labor market enhance contracting possibilities by creating additional possibilities for commitments, and reduce or eliminate the problem of agencies; (2) workers vary in their preferences and in their levels of productivity across different firms; and (3) firms vary in the way the marginal productivity of labor is affected by the business cycle. The modified model presented in this paper assumes, therefore, that efficiency requires workers to switch among firms as business conditions change. Although unreliable information on base employments and layoffs may present limitations on contract, these limitations are generally overcome by the familiar labor market institutions of severance pay, pension and retirement benefits, and a lifetime income curve that rises more swiftly than productivity. The model's accuracy is tested by ten hypotheses about how labor markets function. (JCD)

University of Wisconsin-Madison

Institute for Research on Poverty

Discussion Papers



Barry Nalebuff
Richard Zeckhauser

INVOLUNTARY UNEMPLOYMENT RECONSIDERED: SECOND-BEST CONTRACTING WITH HETERO-GENEOUS FIRMS AND WORKERS

DP #675-81

Involuntary Unemployment Reconsidered: Second-Best Contracting with Heterogeneous Firms and Workers

Barry Nalebuff
Nuffield College
Oxford University

and

Richard Zeckhauser
John F. Kennedy School of Government
Harvard University

October 1981

This research was supported in part by funds granted to the Institute for Research on Poverty by the Department of Health and Human Services pursuant to the terms of the Economic Opportunity Act of 1964. We thank Robert Solow for stimulating our interest in this subject, and for thoughtful guidance along the way. Al Carnesale, Otto Eckstein, David Ellwood, Elizabeth Evanson, Nancy Jackson, Robert Klitgaard, Dutch Leonard, Jim Poterba, Eugene Smolensky, Kip Viscusi, and David Wise provided helpful advice.

ABSTRACT

Why don't wages adjust to clear the labor market? Why can't all workers who are willing to work at the going wages for their skills find a job? A new explanation for involuntary unemployment has been provided recently by implicit contract theory, which shows that despite perfect markets--indeed because of them--wages will not vary over the business cycle. The contracts discussed in the current literature do not, however, capture many of the salient characteristics of real work employment. The performance of markets for employment depends on the opportunities they offer for contracting across time and states of the world. Real world markets, we argue here, operate somewhere between the two extremes of no contracting and perfect contracting--i.e., between the polar cases of spot markets and perfect contingent claims markets. That is, employment relationships usually represent second-best contracts.

Building on implicit contract theory, our formulation takes as its central elements the observed characteristics of labor markets. We assume that the labor force is heterogeneous and mobile, and that there are many types of firms. The latter assumption is required if, as is observed, workers are to switch employers as business conditions change. The overall condition of the economy is uncertain, as are the employment opportunities of a worker and the state of any firm. A critical impediment to labor contracting is that the economic health of the firm may not be verifiable in a manner that can be made the basis for employment relationships. A moral hazard problem thus arises; severance pay effectively overcomes it.

Workers cannot fully commit themselves to firms, because indenturing is not acceptable in our society. Seniority privileges, pension and retirement benefits, and a wage stream that rises faster than productivity (as many do) represent a form of "surety bond" binding workers to firms. Such binding enables firms, in turn, to provide employment contracts that spread risk for workers across good times and bad. Thus, we find that prominent labor market institutions play important roles in facilitating employment contracts.

We develop ten testable propositions from our model, and outline six types of involuntary unemployment--four representing market failures--that flow from it.

Involuntary Unemployment Reconsidered: Second-Best Contracting With Heterogeneous Firms and Workers

INTRODUCTION

Involuntary unemployment arises when individuals who would be willing to work at the going wage for workers with their skills are unable to find employment at that wage. That is, the labor market does not clear. Downward stickiness in the price of labor is usually held responsible for this condition. Such stickiness has been attributed to (1) employed workers' efforts to defend wage differentials (Keynes, 1936), (2) fair play and the fulfillment of social norms, (3) the minimum wage, unemployment insurance, and their by-products, (4) collective bargaining between unions and firms (Solow and McDonald, forthcoming), and (5) implicit contract mechanisms[1] that allow firms to provide insurance to more risk-averse workers by leveling their wages over the collection of states--that is, the different conditions--in which they are employed.[2]

Implicit contract models (described below), by highlighting the type of long-term employment relationships we observe most commonly in the real world, represent a significant advance over the conception of employment as a series of independent transactions on spot markets, which implies that there is no contracting over time. Most formulations of the model to date, however, fail to take account of the factors that make it difficult to enforce wage and employment contracts across states of the world. Such limitations lead to market failures that are a source of involuntary unemployment. Moreover, these limitations create a need for labor market

institutions that enhance contracting possibilities. The truth about employment, we would argue, lies somewhere between the two extremes of no contracting and perfect contracting across time and states of the world, i.e., between the polar cases of spot markets and perfect contingent claims markets. Employment relationships thus represent "second-best contracting." Recognition of this truth should help build employment contract models that deepen our understanding of why labor markets do not clear.

We attempt to develop a theory that recognizes several real-world circumstances ignored in the traditional implicit contract model: (a) institutional characteristics of the labor market--severance pay, pension and retirement benefits, and a rising lifetime income path for employment within a single firm--enhance contracting possibilities by creating additional possibilities for commitments and reduce or eliminate the agency problem; (b) workers are heterogeneous; they vary both in their preferences and in their productivities across different firms; (c) firms are heterogeneous; in particular they vary in the way the marginal productivity of their labor is affected by the business cycle.

In addition to providing new explanations of involuntary employment and the downward stickiness of wages, a theory of second-best contracting that builds on the implicit contract model can suggest a series of intuitive, empirically testable propositions. Given the specificity of these propositions, they are also capable of being disproved. Thus they permit more strenuous tests of the implicit contract concept than does the traditional version of the theory, which was developed primarily to explain the already well-known phenomenon of downward wage stickiness.

Because we would like these ideas to reach a broad audience, our analysis is more in words than symbols. In future studies, we hope that we and others will provide more rigorous presentations as well as develop empirical assessments of our hypotheses and other hypotheses that flow from this formulation.

THE IMPLICIT CONTRACT MODEL OF LABOR MARKETS

The standard approach to implicit contract theory assumes that a single type of worker who is risk-averse is hired by a single firm that is risk-neutral, and that the worker if laid off has one alternative--or as we term it, one fallback opportunity--which is usually unemployment (and its associated benefits). The firm pays the worker both when employed and when laid off (through unemployment insurance). The hope would be that an employment contract could be drawn that would insure the worker against the risks associated with changing employment and business conditions, yet always place the worker in employment where the value of his product is greatest.

To simplify our presentation, we shall generally use a one-period formulation, where at the time the labor contract is drawn there is uncertainty about the state of the world. The worker must consume his whole wage in a period.[3] There are either fixed costs of getting to work, institutionally fixed work weeks, or increasing returns to leisure, all of which imply indivisibilities in the supply of labor. Define

W_e as salary when employed,

W_u as salary when unemployed,

ℓ as labor, $\bar{\ell}$ when employed

0 when unemployed, and

$U(W,\ell)$ as worker's utility.

In order to attract any workers, the firm must provide its employees with an expected utility of at least $\bar{U}$, a level that is determined in the competitive market. To maximize a worker's utility for a given expected wage bill, and thereby minimizing the cost of generating $\bar{U}$, a firm offers a constant salary of W_e during the states in which the worker is employed and W_u when he is unemployed.[4] To achieve efficiency, the salary payments should be like an insurance contract. Thus, the laborer's marginal utility of income should be the same when he is employed as when he is unemployed.

$$\frac{\partial U(W_e,\bar{\ell})}{\partial W} = \frac{\partial U(W_u,0)}{\partial W}$$

Say the worker's utility function is of the form $U(W,\ell) = V(W + \alpha\ell)$, where α is the implicit dollar price of work effort. Then the optimal contract would have $W_e = W_u + \alpha\ell$. Workers would be unemployed only if the value of their output is less than their disutility of work. This unemployment is not involuntary in the economist's sense, because only workers who cannot produce output of sufficient value to compensate for their disutility of labor would choose to be unemployed. Moreover, workers are indifferent toward working at W_e or taking leisure at W_u.

By contrast, if the income and leisure arguments of the utility function are separable and additive, the optimal contract would have the wage when unemployed equal to that of the wage when employed.

Critique of the Standard Model

The standard model of implicit contracts has been criticized on a variety of grounds; three seem central.

First, the model does not appear to describe the qualitative aspects of many important labor markets in which involuntary unemployment is a problem. Although it does assign a role to identifiable institutions, such as unions or the reputations of firms, which provide some of the elements needed for a contracting mechanism,[5] many of the central tenets of implicit contract theory would be almost unrecognizable to those who should be most concerned, namely workers and firms.[6] For example, the theory assumes that workers and firms draw lifetime employment contracts, yet labor mobility among firms is a key feature of most modern economies.[7] The standard model also assumes the existence of contracting possibilities that are not readily available. In the real world, a worker's difficulty in discerning the general state of the economy, the specific condition of his employer's firm, and even his own ability creates an "agency problem," if we regard the firm as the agent for the worker. When the value of an employee to the firm is less than his wage, the firm will have the incentive to fire him; to get around a contractual obligation, it might distort information and say that the value of labor to the firm was lower than it truly was.

Moreover, contrary to the assumptions of the standard model, real-world labor forces are highly varied: different workers are laid off at different times, others switch to new occupations. Recognizing this heterogeneity, public policy toward employment often attempts to change the characteristics of frequently unemployed workers to make them more like others.

The heterogeneity of firms is a further salient but underrecognized aspect of markets for employment. Firms may differ in their sensitivity to the business cycle and in the relative productivities of different types of labor. Even if workers were homogeneous, heterogeneity among firms would lead to a major departure from the traditional implicit contract theory: Across different stages of the business cycle, workers should be--i.e., with optimal contracting would choose to be--employed at different firms.

Second, the standard model provides only a limited explanation of "involuntary" unemployment. (The notable exception is Grossman and Hart, 1981; see footnote 13.) We shall describe several different types of involuntary unemployment that arise under our formulations. Some represent the true market imperfections that economists require to consider the condition involuntary. Other definitions are less restrictive and merely capture what the public or unemployed workers might mean by involuntary unemployment.

Third, the standard theory has not been verified by reference to empirically testable propositions. Since there may be many possible

ways to explain involuntary unemployment, we should seek more refined and subtle propositions that might enable us to distinguish among alternative theories. In this paper we identify ten testable propositions, and we compare the predictions of our model with the predictions derived from models that have (1) only spot markets for labor, and (2) perfect contingent claims markets for employment.

CONTRACTING GIVEN HETEROGENEITY OF WORKERS AND FIRMS

We start by expanding the standard model to include a heterogeneous labor force that has employment opportunities among a range of firms. In the earlier models, firms were constrained to offer workers contracts yielding an expected utility of at least $\bar{U}$. The level $\bar{U}$ was determined by the competitive assumption of zero profits. With a heterogeneous supply of workers, each employee might reach a different level of utility with each firm. In a world of perfect contracting, risk-neutral firms will in effect let their employees design their own optimal contracts subject to a zero expected profit constraint for that type of worker. A firm that did not follow this policy would lose all its workers to a competitor who could offer them a higher utility without losing any profits.[8]

Implicit contract models traditionally assume that workers and firms draw up contracts that commit them to a lifetime employment relationship, usually even covering periods of unemployment. It is not possible to explain labor mobility or even transitional unemployment within such a

model. Worker heterogeneity does not provide an explanation, at best it would generate different layoff experiences. Once we allow for heterogeneity of firms, however, labor mobility becomes not only explainable, but essential to achieve efficient outcomes. If firms are heterogeneous, the firm at which the value of a worker's product is highest will depend on the state of the world.

We shall refer to the value of a worker's product at a firm in a particular state as his "value added." Value added is computed by multiplying the worker's physical productivity times the unit value of the output he produces. A priori, it is not possible to say which of these two elements produces greater variability in a worker's value added. The physical product of construction workers varies with the weather. The value added of an auto worker, by contrast, is likely to depend predominantly on the strength of demand for the car he produces.

If firms are differentially affected by business conditions or seasonalities, we might expect that some--say auto manufacturers or construction firms--would use layoffs regularly, others hardly at all. Some activities, such as own-home repair, will pick up considerable numbers of workers in poor times. In the analysis below, we shall be particularly interested in the process of reallocating workers among firms across variations in business conditions. How successful this reallocation process turns out to be will depend on the characteristics of the labor market.

Most microeconomic analyses of labor markets fall into one of two polar cases: (1) spot markets that clear on a period-by-period basis,

Table 1. Our Analysis of Labor Markets Compared to the Two Polar Models

Elements in the Analysis	Polar Model 1: Spot Markets Only	Our Approach: Second-Best Contracting	Polar Model 2: Perfect Contingent Claims Markets
Characteristics of Markets	New employment market occurs in each period after information (e.g., business conditions) becomes known.	Potential for (a) contingent contracting for wages over collections of states, or (b) long-term contracting between worker and firm.	Workers contract with a number of firms for contingent employment prior to each period. Sign-up bonuses exist.
Information Structure	Given competition information asymmetries not always relevant.	Information asymmetries possible and relevant.	No information asymmetries.
Locus of Employment	Workers take highest value-added employment.	In interest of risk spreading, workers may not take highest value-added employment.	Workers take highest value-added employment.
Enforcement of Contracts	No commitment necessary.	Commitment to contract through (a) reputations--firms and workers; (b) severance pay--paid by firms when the dismiss workers; (c) surety bonds (e.g., seniority and retirement benefits, wage rising faster than productivity) sacrificed by workers when they leave firm.	Labor contracts enforceable without cost.
Properties of Outcome	Productive efficiency; no risk spreading.	Sacrifices in productive efficiency accepted to promote risk spreading.	Productive efficiency; perfect risk spreading.

and (2) perfect contingent-claims markets for employment. Our analysis addresses the class of somewhat cloudier situations where workers and firms can contract with each other either over continuous blocks of employment or before business conditions become known. However, such contracting is hampered in several ways: by imperfect commitment mechanisms, by the lack of bonus arrangements in signing employment contracts, and because workers and firms may not have the same information. Table 1 compares our approach with its two polar counterparts.

The traditional implicit contract formulation employs a perfect contingent claims market. Where there is a single fallback opportunity, the sign-up bonus is paid as unemployment insurance if unemployed, or if employed, as a component of the wage.

In our models, unless stated clearly to the contrary, we make the following assumptions:

1. Firms know the state of nature; workers may or may not know the firm's economic condition.
2. Although ex ante contracting may be feasible, no bonuses or penalties can be paid before a worker comes to a firm.
3. The worker's marginal utility function for money is state-independent.[9]

In our formulation, firms contract with workers over a set of states. The wage they offer depends on both the identity of the worker and the states covered. (Thus the $\bar{U}$ of implicit contract theory would have a value depending on two indices.) We further define:

$W_{ij}(s)$ as the value added of a type i worker at firm j in state s,

p_s as the probability of state s.

s^*_{ij} as the collection of states, s, in which worker i finds it optimal to work for firm j, and

$\bar{W}_{ij}$ as the average (or expected) value added of worker i when he is with firm j.

Thus we have

$$\bar{W}_{ij} = [\sum_{s^*_{ij}} p_s W_{ij}(s)]/[\sum_{s^*_{ij}} p_s].$$

Given the wage-averaging feature, optimization at the margin may not lead to a globally optimum solution. Although it is very difficult to solve the nonlinear programming problem for the optimal states of employment of a worker with a firm, s^*_{ij}, we can characterize many features of the solution.

We follow Akerlof and Miyazaki's (1980) wage bill argument in asserting that the worker will seek a constant wage over the range of states in which he remains with a firm.[10] Risk-averse individuals will always prefer to receive the expected value of a random variable. The zero expected profit constraint implies that worker i can command $\bar{W}_{ij}$ while working for firm j.[11]

The implicit contract mechanism can provide insurance to workers for their varying productivities across states of nature while employed at a single firm. It does not provide insurance across firms that would allow a worker to receive the value of his average productivity between two firms, or even between interrupted periods of employment with a single firm. The

optimization problem for workers has been somewhat simplified. Worker i chooses s_{ij}^*, the range of states to work for firm j, and is paid $\bar{W}_{ij}$ during that time. Once the heterogeneity of firms is recognized, the implicit contract model becomes consistent with the considerable job switching and labor mobility that we observe in the real world. (Hall, 1981, concluded that in the United States, the average job tenure was eight years.)

An Example of an Optimal Contract: The Base Case Model

There is one worker who can be employed by either of two firms. His value added is

$$W_{11}(s) = 12$$
$$W_{12}(s) = s,$$

where s is uniformly distributed on [0,20]. That is, the worker has a constant fallback wage of 12. His value added to the second firm varies uniformly between 0 and 20. If the worker always went to the firm where his product was higher, we would have

$$s_{11}^* = \{s: \ s < 12\}$$
$$s_{12}^* = \{s: \ s \geq 12\}$$
$$\bar{W}_{11} = 12 \text{ and } \bar{W}_{12} = 16.$$

Then, under optimal contract he would earn \$16 when employed with the second firm and rely on his fallback opportunities to earn \$12 when his value added to the second firm was less than \$12.

Switching employment at \$12 serves the objective of productive efficiency, but neglects risk-spreading goals in this second-best

situation where funds cannot be transferred between employment at firm 1 and firm 2. Optimality requires a risk-averse worker to stay with the second firm even below the point at which he could earn $12 with the first firm. This would allow him a greater opportunity to average his income. Let the switching point between firms 1 and 2 be X. Then a worker would maximize as follows:

$$EU = \frac{X}{20} U(12) + (1 - \frac{X}{20}) U(\frac{20 + X}{2}).$$

His first-order condition is

$$\frac{\partial EU}{\partial X} = \frac{1}{20} [U(12) - U(\frac{20 + X}{2})] + (1 - \frac{X}{20}) U'(\frac{20 + X}{2}) \frac{1}{2} = 0.$$

At X = 12, $\partial EU/\partial X < 0$ if U is concave. Productive efficiency is sacrificed so as to spread risk across as wide a set of states as possible. When the utility function is logarithmic, for example, the optimal switching point is at X = 11.60.

Normally, we think that implicit contracts decrease labor mobility. This intuition can be misleading, as evidenced by our example. With spot markets, a worker at the second firm has only a .4 chance of staying, while an employee of the first firm remains with probability .6 (s < 12). Under an implicit contract, the worker will stay at the second firm until his value added there is less than 11.60, so his chance of moving is .42. The expected transition probability = 2 x .42 x .58 = .49, which is now slightly higher than it was with spot markets: there is greater mobility of labor. This increased mobility helps spread risk, but is inefficient from the standpoint of production.

IMPEDIMENTS TO CONTRACT

The refined contracting model we have outlined still does not capture many important features of the way labor markets actually work. The implicit contract formulation it employs assumes that the information on the basis of which contracts are drawn, namely the state of nature, is publicly available to both workers and firms. It requires as well that firms and workers be able to assure each other that they will stand by their commitments. Neither of these assumptions is always satisfied in practice.

Continuous Blocks of Employment

The contracts we have thus far considered permit workers to draw contracts with firms before they are employed and the state of nature is known. A more realistic framework would permit contracting only with a worker's present employer. Starting at firm 1 (where the worker is always worth 12), there are two advantages in switching to firm 2 ($W_{12}(s) = s$). First, he will then be able to contract for future periods with firm 2. Second, his surplus (or deficit) wage in the transition period can be averaged over the expected lifetime of the contract. These two classes of benefits must be balanced against the worker's lower output value when $s < 12$.

A worker who enters firm 2 when worth s and contracts to stay at firm 2 provided future states, $s > x(\hat{s})$, has an expected employment span with the firm of $\frac{20}{x(\hat{s})}$ periods. During that time he is paid what he is worth on average,

$$C(\hat{s}) = \frac{x(\hat{s})}{20} \cdot \hat{s} + \left[1 - \frac{x(\hat{s})}{20}\right] \cdot \left[\frac{20 + x(\hat{s})}{2}\right].$$

The optimal state in which to exit from firm 2, $x(\hat{s})$, will in general depend on the entry state, $\hat{s}$. In Appendix 2 we prove first that $x'(\hat{s}) \leq 0$, the greater the worker's productivity when he enters firm 2, the greater the span of states over which he will remain with firm 2; and second that at the smallest salary at which it is still optimal to remain with firm 2, denoted $\underline{\hat{s}}$, the exit state will equal the entry state, $x(\underline{\hat{s}}) = \underline{\hat{s}}$. For the parameters defined there, the optimal switching strategy is shown in Figure 1. When the first period is not obscured by "the veil of ignorance," the contracts negotiated between workers and firms depend on the worker's value to the firm when he is hired.

Firm-Specific Information Not Shared with Workers

When workers are unable to monitor fully the state of nature, moral hazard produces a major difficulty in the implicit contract formulation. Perhaps workers can judge the overall well-being of the economy or of industries, but it is much more difficult to measure the well-being of particular firms. Because a worker's salary at a firm is the average over states of nature of his value added with that firm, there will be some state(s) in which his value added is less than his wage. In such states, a firm would have an incentive to tell the worker that the current state of nature is not in s_{ij}^*, i.e., s is below $f(\hat{s})$ in the model just considered, and that he should seek alternative employment.[12] Since a worker could not expect to be employed if his salary exceeds his value

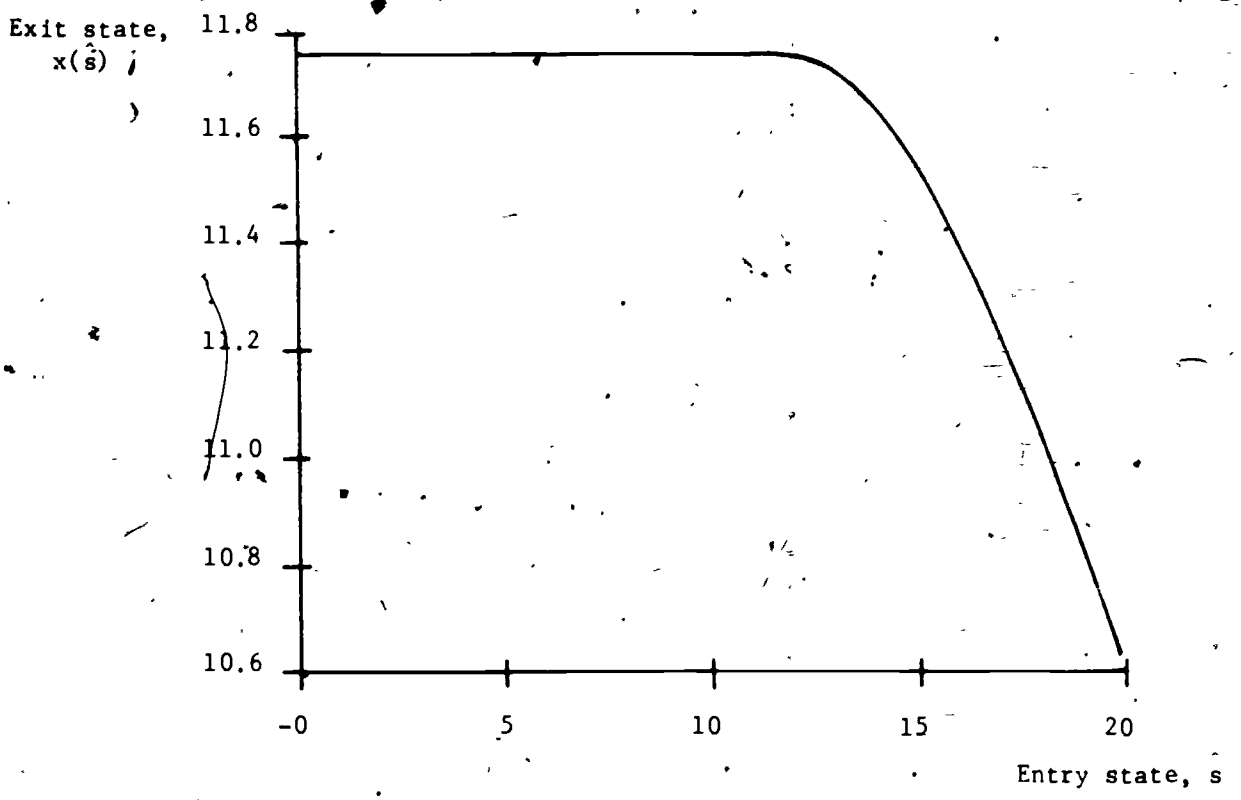


Figure 1. Optimal Switching Strategy

added, the process will tend to unravel from the bottom[13] and all labor will be hired in spot markets.

Workers would like to design a contract that gives firms incentives to keep employees even when their wage exceeds their value added,[14] thereby allowing for some useful risk spreading. The solution to this moral hazard problem is to make the firm give the laid-off worker a lump sum financial payment, the quid pro being a sacrifice in salary when employed.[15]

Severance pay. We shall refer to such lump-sum payment as severance pay, recognizing that such payment could be made through a range of institutional structures. Severance pay is paid only if a firm chooses to fire one of its workers, not if the worker leaves for other employment or any other reason. We return to the simpler framework of the Base Case Model to show how severance pay operates.

Even if contract enforcement were not a problem, say because information flowed fully and freely, severance pay would be worthwhile solely as a risk-spreading mechanism. In the Base Case Model, rather than receiving a wage of \$16 for 40 percent of the time and \$12 the other 60 percent, the worker would be able to receive \$13.60 all of the time. The worker would simply negotiate a contract with firm 2 to work there at a wage of \$13.60 whenever $s \geq 12$, and to receive severance pay of \$1.60 when he was laid off, i.e., $s < 12$.

Salary in period if employed at firm 2	\$13.60	$s \geq 12$
Total renumeration in period if laid off	\$12 + 1.60 = \$13.60	$s < 12$

Happily, the use of this risk-spreading instrument solves the imperfect contracting problem as well. The firm will have an incentive to behave in accordance with its contractual obligations. Although a worker may earn $13.60 when he is only worth $13, a firm would lose $1.60 in severance pay by dismissing him, hardly worthwhile to save $0.60. It would choose to employ its workers provided they are worth at least $12 ($s \geq 12$), and thus productive efficiency is achieved.

Severance pay eliminates the employer's moral hazard temptation, and equalizes workers' pay across the entire range of possible states of nature, including ones in which it is appropriate for the workers to move from firm to firm.[16]

A converse information problem could arise if workers have relevant information not available to the firm. Under an optimal contract, workers would change employers only when their earnings gain exceeded some threshold, but would remain at their original firm for a range of states in which they could earn somewhat more elsewhere. Presumably, workers would have an incentive to leave for any salary gain whatsoever, asserting falsely that it lay above the cutoff level.

Prohibition against Indenturing Workers

The major problem in securing the worker's efforts at a firm may be in enforcement of the contract. Our society does not in general permit workers to indenture themselves to firms, even at a positive wage. Thus even if information were fully shared, a serious problem arises when

times are good and workers are earning less than their product. In these situations, workers may realize that they are affording surplus value and seek other employment. (Even if workers do not know their own value, other head-hunting firms would try to steal "underpaid" workers.)

A laborer who is unable to commit himself to a firm when his product is greater than his wage is also unable to ask for any insurance against times when his product is less than his wage (Azariadis, 1976; Grossman, 1978). If a worker is to average his income over good times and bad, there must be some mechanism tying the worker to the firm during periods of surplus: severance pay plays the role of committing the firm to the worker. But it is very difficult to create an enforceable financial penalty to prevent a worker who is underpaid (even if only temporarily) from quitting.

To commit themselves to a firm, workers must rely on more subtle and less effective mechanisms. A rising income path that is steeper than accompanying gains in productivity, seniority privileges, and retirement plans can be seen as a sort of "surety bond"--a mechanism that rationally commits the worker to stay with a firm even when his value exceeds his pay.[17] This surety bond represents a worker's counterpart to severance pay. But it is relatively more important, since a worker's reputation contains much less information than a firm's, and the need to maintain a good reputation (e.g., he is not a job hopper) is not likely to be as powerful an inducement to honor an implicit contract to workers as it is to firms.

The worker will only leave if his product is enough greater than his wage to justify the loss in seniority and retirement benefits, or

in the "promised" earnings in excess of productivity.[18] Again, the worker must make a tradeoff between deferring more income into the later periods and accepting a lower mean salary early on. Depending on his risk aversion and fallback opportunities, a worker may choose to defer only a small part of his earnings; in this case he will have only a limited ability to negotiate risk-spreading contracts.

In our earlier example, a hypothetical worker earned $13.60 when employed by firm 2 throughout a range of states in which his value added fluctuated between $12 and $20. This worker would have to put aside $6.40 as a surety bond to convince the firm that he would stay with them even if his product were to be as high as $20. If the worker voluntarily leaves the firm he would forfeit the surety bond. Although another firm with the same characteristics as firm 2 might offer him $19 when s = 19, he knows that he would lose $6.40 by quitting his $13.60 job, and he therefore remains.[19]

The introduction of a surety bond lowers the worker's utility in two ways. First, it disrupts his smooth salary pattern as income earned early in his career must be put aside into retirement plans (and seniority privileges). Second, it may be appropriate--risk aversion and efficiency considered--for a worker to switch to a better-paying job.[20] By doing so he would forfeit his retirement benefits and thus indirectly subsidize the other employees.[21] These considerations may stop a worker from putting aside funds sufficient to overcome his moral hazard problem. Since he is no longer willing to commit himself to a firm in all good states, he must

also accept layoffs in a greater number of bad states.[22] Severance pay represents a further difficulty, because it lowers an employee's salary when working. This exacerbates the need for a surety bond (seniority privileges, rising income gradient, etc.) to keep a worker from quitting.

It takes time to build up a fund for retirement or other seniority privileges that can be used as a surety bond. New entrants to the labor force and young workers especially will have difficulty in this area, which is surely one reason for their observed high turnover rates.[23] Unfortunately, the problem is self-reinforcing. Firms that are unable to benefit from a worker's surplus value in prosperous periods cannot afford to offer any insurance (and severance pay) during more difficult periods.

A worker's tradeoff between salary and job security is affected not only by skill and comparative advantage, but also by risk aversion. The mechanisms that bind workers to firms and firms to workers are not without cost. In this second-best situation, workers must choose between being tied to a firm through large seniority benefits, and being able to take advantage of higher-paying, more productive jobs with greater employment variability.

While committing a worker to his firm, seniority and retirement benefits may also induce him to work harder. In a model studied by Lazear (1979), workers do not flag in their effort, since that would increase their chance of being fired and thus forfeiting their forced saving (both retirement benefits and the privileges and higher wages that come with seniority). Firms, of course, would like to fire workers

once they are obliged to begin paying back the worker's savings, but are prevented from doing so by the threat of union reprisals. (They should also be concerned about loss of reputation as a good employer.)

In our formulation, we allow the firm to retain the value of an employee's surety bond only if the worker quits. If a firm wishes to fire a highly tenured worker, it must pay him both the regular severance pay and all his earned seniority benefits. Many a lawsuit and not a few regulatory policies have been addressed to this class of issues.

The importance of seniority as a part of the wage contract is confirmed by the recent empirical investigations of Medoff and Abraham (1981), which show that (a) seniority rather than merit is the main criterion for promotion even in nonunion ships, and is particularly important in unionized ones,[24] and (b) the productivity path with seniority is almost flat, while the wage trajectory is rather steep, so that senior workers are being paid more than their value added and junior workers less. Seniority benefits are perhaps too great to be viewed solely as a form of surety bond.[25] Yet these benefits are an essential element of many labor contracts, and are consistent with the need for a mechanism that rationally commits a worker to a firm.

In the following section, we continue an intuitive discussion, revolving around a series of testable hypotheses about optimal contracts. The optimal contracts will depend on a worker's utility function (risk aversion), his value to the employer (absolute advantage), and the opportunities on which he can fall back (comparative advantage).

The term comparative advantage is very broad in scope. It is used here to relate what a worker earns with one firm to what he will earn at another, W_{i1}/W_{i2}, across different states of the world. One of the firms may represent leisure, unemployment, or work in a secondary labor market. For many of the testable hypotheses that follow, it is convenient to consider the special case in which there are only two firms. As in our earlier example, we have one firm at which each worker has an uncertain value, $W_{i2}(s)$, and a second firm that offers each worker a constant fallback opportunity, $W_{i1}(s) = W_{i1}^0$. The ratio of $W_{i2}(s)/W_{i1}^0$ is the comparative advantage of a worker of type i for firm 2 in state s relative to the fallback state.[26] The consideration of comparative advantage in determining each worker's optimal employment contracts over states and firms is of central importance in the analysis that follows.

PROPOSITIONS TO TEST OUR FORMULATION

In this section we put forward ten propositions about how labor markets will function under second-best contracting. These can be tested empirically, to gauge the accuracy of our model. For comparison, in each case we note first the predictions that would flow from the two major alternative models. They assume that employment transactions take place respectively in (1) spot markets, which allows for no risk spreading across periods and states, and (2) perfect contingent claims markets, which is the implicit-contract formulation in its pure form. An explication of the second-best contracting result follows each prediction.

Proposition A: Wages and Skill Levels

Spot Markets: Two workers equally skilled in a common job will earn the same amount at that job.

Contingent Claims Markets: A worker's salary is constant and reflects his skill averaged over states of nature. Two workers equally skilled in a common job will earn different amounts if their skill levels differ for other employment situations.

Second-best Contracting: Two workers who are equally skilled in a common job will voluntarily contract to earn different amounts at that job despite perfect competition, because of differences in their opportunities in other states of the world. Specifically, the one with a higher fallback wage will have a lower comparative advantage at the present job, will require less job security, and will receive a higher wage. Further, a worker who earns more than another at one firm may earn, at another firm, less than the former coworker.

This suggests as well that, other things being equal, the distribution of workers earning high salaries will be skewed toward those with high fallback wages and thus a lower comparative advantage for the present job.[27] We informally think of transition costs as lowering a worker's salary at all firms other than his current employer. High transition costs (a complication to be dealt with in future work) will cause a worker to have a higher comparative advantage in his present job and thus he will seek jobs with greater employment security and correspondingly lower wages. Comparative advantage may be difficult to measure or observe; still, it can provide an

explanation for different salaries and employment security among workers who are similarly skilled in a job.

Proposition B: Layoffs and Skill Levels

Spot Markets and Contingent Claims Markets: There are no lay-offs representing involuntary unemployment.

Second-best Contracting: The first workers laid off during bad times, i.e., when the worker's value added declines, may not be those who are least capable at the present job.

If implicit contracts define the employment relationship, workers with lower comparative advantage will be the first to be dismissed. If the greater adaptability and flexibility of more capable workers makes them more valuable to a different employer, they may be laid off early. When a high technology market declines (hand calculators, for instance), it may be the more outstanding engineers who move to new fields.[28] Similarly, if A values his leisure twice as much as B, his comparative advantage will be lower even if his productivity is 150 percent of B's. A will be laid off first.

Proposition C: Wages as Indicators of Quality

Spot Markets: Individuals who earn more than others in some states may earn less than these same coworkers in other states.

Contingent Claims Markets: Workers' wages are always in strict rank order.

Second-best Contracting: The concept of labor quality is ambiguous. Individuals who earn more than others in some states may earn less than these same coworkers in other states. In particular, less risk-averse workers will have greater variability of wages.

What is meant by the statement: "Holly is a higher-quality worker than Janet"? Unless $W_{ij}(s)$ is larger for Holly than for Janet, both over all states, s, and over all firms, j, there is room for ambiguity. Even if this very restrictive test is met, there is no presumption that the higher-quality worker will necessarily earn more than the other.

The Value Added of the Workers

Holly/Janet		Firm 1	Firm 2	
s	1	11/10	9/8	p(s = 1) = .5
	2	1/0	4/3.5	p(s = 2) = .5

Holly could be sufficiently risk-averse to choose firm 2 in both states while a risk-netural Janet would work for firm 1 in s = 1 and firm 2 in s = 2. Janet's average salary would be 6.75, which is greater than Holly's constant wage of 6.5. It seems very difficult to distill an intuitive concept of worker quality from a general model that uses both comparative advantage and risk aversion to determine optimal employment contracts. (Further complications arise because the value added of a worker also depends on the price of the output he produces, a variable that puts additional noise into the system.)

Proposition D: Choice of Employer

Spot Markets and Contingent Claims Markets: Workers always choose the firm at which their value added is highest.

Second-best Contracting: Workers will not always choose to work at the firm where their value added is greatest. The employment package of another firm may be preferred because it spans a broader set of states of nature.

A laborer may choose to work for a particular firm even though he might always be more productive elsewhere.

Value Added of Holly

s \ Firm	A	B	C	
1	6	7	0	p(s = 1) = .5
2	30	0	31	p(s = 2) = .5

In this example, Holly might choose to work always at firm A so as to earn a constant salary of 18 rather than produce more in each state but face a varying income of 31 or 7.[29]

Proposition E: Simultaneous Hiring and Firing

Spot Markets and Contingent Claims Markets: Firms are usually treated as homogeneous. Only if firms are heterogeneous will firing and hiring occur in the same job category.

Second-best Contracting: As part of the natural process of the efficient reallocation of workers across firms in different states of nature, we should see both hiring and firing in the same job category when conditions change downward. Conversely, when conditions improve, we should observe some workers leaving their jobs (to move upwards) and other new people being hired to fill their places.

Consider faculty appointments in history departments. A shortage of opportunities for historians will hurt young graduates seeking their first jobs, since more mature scholars will also be competing for the few positions available. The available mature scholars are likely to prevail--perhaps not because they are of higher quality than the young historians, but because they have invested more in their academic careers and thus have a higher comparative advantage at (are more desperate for) the professorial jobs.

Proposition F: Firm Productivity and the Business Cycle

Spot Markets: A firm's productivity measured in value added/wage will be constant over the business cycle.

Contingent Claims Markets: No variability in gross wages (period wage plus bonus) across the business cycle; hence, this productivity measure will vary significantly.

Second-best Contracting: Quite apart from any changes in the composition of the work force, a firm's productivity measured in terms of value added/wage should fall during bad times and rise

during good times because of the risk-spreading features of optimal contracts. It will vary less than it would with perfect contingent claims markets.

The price of outputs relative to wages will tend to fall during a recession and many workers will be earning more than their value added. Part of the negotiated wage contract involves insuring workers against bad times. The fact that these insurance contracts are paying off during recessions contributes to the significant fall in the measured value added/wage.

Value Added of Holly

		Firm 1	
s =	1	30	p(s = 1) = .5
	2	10	p(s = 2) = .5

In this example, Holly would earn a constant wage of 20. When conditions fall from state 1 to 2, her value added/wage would decline from 1.5 to .5. Were her wage adjusted, there would be no decline in productivity measured as value added/wage.

Layoffs will mitigate or exacerbate the firm's decline in productivity depending on whether the workers with the smallest comparative advantage, i.e., those laid off first, also have the lowest absolute advantage.

Usually we would expect both physical productivity/worker and value of output to be positively correlated with business conditions, which

implies that both will fall during bad times, but less precipitously than will value added/worker.

Proposition F': Profits and the Business Cycle

Spot Markets: Variations in measured profits represent only changes in returns to capital.

Contingent Claims Markets: Profits rise dramatically during boom times.

Second-best Contracting: Measured profits will rise during boom times, since the tendency to pay average wages will increase productivity (value added/wage). They will rise less dramatically than with perfect contingent claims markets.

This is primarily a corollary to our previous proposition, since we observe that the largest factor affecting profits is a change in productivity. Interestingly, if different industries and firms benefit differentially from a boom, the rise may level off or even diminish, since workers will appropriately reallocate themselves to new firms.

Proposition G: Worker Reallocation Affecting Productivity

Spot Markets and Contingent Claims Markets: Reallocation of workers over the business cycle plays no risk-spreading function; it serves solely to achieve maximum value added in each state.

Second-best Contracting: The reallocation of workers will mitigate the overall economy's decline in productivity during a recession. Because of biases in measurement, the reported fall in productivity

may be misstated. Mobility may be greater or less than with each of the alternative models.

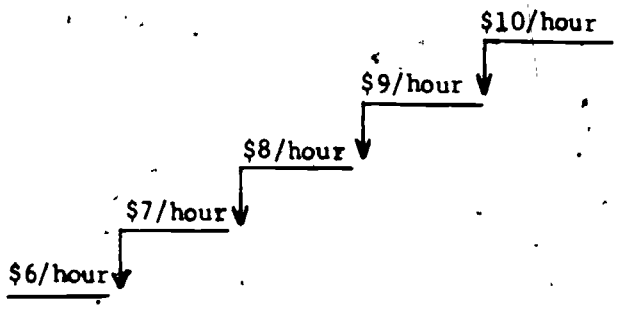
As workers switch firms during a recession, the natural fall in productivity will be dampened by the fact that the workers who are laid off first are the ones with the lowest comparative advantage. They will earn relatively the most in other modes of employment, including leisure or underground activities.

Another appropriate measure of productivity is

$$\frac{\text{total value of output}}{\text{number of workers}},$$

where total value of output is the sum of value added for the employed and value of leisure for the unemployed. Productivity measured in this way may suffer from a number of biases. Ideally, such an index would measure the value of leisure or nonmarket work activities such as home repair for a constant sample of workers. In practice, holding the sample constant is almost impossible, and computations of productivity relate solely to employed individuals. This introduces at least two biases: (1) during a recession, to the extent that more low-productivity workers are forced to choose activities not metered through the market, there will be a bias to understate the fall in productivity; (2) if workers are measured in units standardized for productivity--in an attempt to deal with bias (1)--there will be a bias to overstate productivity loss, since the "lost workers" will be the ones with relatively higher productivity elsewhere.

There is a potential difficulty in capturing the new composition of the general economy as the relative importance of particular industries shifts during a recession. It is possible to construct an example in which the productivity of each firm rises from a shuffling of workers but the economy's overall productivity falls.



In the diagram above, each firm lays off its worker with the smallest comparative advantage and hires the worker laid off from the firm above it.[30] After such a shuffle, the productivity of each firm could rise if the newly hired worker is of greater value than the recently dismissed worker. Yet the economy's productivity may have fallen. The relative share in production of the top (and most productive) firm has declined, that of the least productive firm has increased. Unless measures of productivity accurately compute both the change in each industry's productivity and the new relative importance of each industry, there will be a bias toward higher reported productivities during recessions (and lower reported productivities during booms) due to the escalator effect described above. (For a related difficulty in measuring the consumer price index see Gordon, 1981.)

Proposition H: Firm Layoffs Due to Relative Performance

Spot Markets and Contingent Claims Markets: With heterogeneous firms, firm performance relative to the economy is important to the labor allocation process.

Second-best Contracting: A firm's layoffs will depend on how it does relative to the economy. When a firm does poorly but the general economy remains stable, we should observe more layoffs than when the firm does poorly and the economy is also depressed. Given risk-spreading considerations, departures from firms are less precipitous than with the two alternative models.

Fallback wages are much more sensitive to the general condition of the economy than to the conditions of any one firm. A worker's comparative advantage may remain constant if both his employer and the economy move down together. However, comparative advantages will surely fall for workers at a firm that is doing worse than the general economy. Workers may be willing to accept lower wages to remain employed when everything is depressed since their fallback opportunities might then be significantly fewer.[31] But when the general economy is healthy, a worker might prefer to switch firms if his employer is doing relatively badly. (The same principle applies to shifting out of professions or industries.) This also suggests that it might be advantageous to work for a firm whose business is counter-cyclical--that is, does poorly when the economy does well, and vice versa.

Proposition I: Order of Unemployment and Reemployment

Spot Markets and Contingent Claims Markets: There are no layoffs, even with heterogeneous firms. Departure from and return to a firm

is a last-in-first-out (LIFO) process.

Second-best Contracting: Employment for an individual firm or for the economy as a whole is essentially a LIFO process.

The workers with the relatively best alternative opportunities will be the first ones chosen to be laid off. The workers who are laid off last are the ones least equipped to handle the layoff, in the sense that their comparative productivity is highest. Thus they will also be the first ones to be rehired. Lazear (1980) relates a worker's elasticity of labor supply to his second-best alternative use of time. He concludes: "This suggests that the young workers (whose alternative may be school) are the first to be laid off in low season. As such, layoff by reverse seniority can be viewed as part of the efficient contract." To the extent that this observation is valid, it might cause us to rethink some of our conventional attitudes about where to direct training, retraining, and employment policy.

Proposition J: [Interstitial] Unemployment

Spot Markets and Contingent Claims Markets: There are no gaps between episodes of employment. Hence, there are no gaps between episodes of employment. Hence, there is no interstitial unemployment.

Second-best Contracting: Unemployment should rise when the general economy is in a state of flux since a larger number of workers will find it appropriate to switch jobs and will have to accept some expected interstitial unemployment to make the switch.

Unemployment depends on the variability of the economy and not just on its absolute level. As an economy shifts, many workers' comparative advantages change and they become able to improve their productivity by switching jobs.

The problem of unemployment and reemployment caused by a period of economic transition is exacerbated by workers' loss of seniority (i.e., promise of wage in excess of productivity) and retirement or severance benefits when they change to more productive jobs.[32] One important benefit of policies that lead to a more stable economy would be to reduce those potentially high transition costs.

CONCEPTIONS OF INVOLUNTARY UNEMPLOYMENT

We are now in a position to discuss what we consider to be six types of involuntary unemployment. The first two cases are not true market failures, although they are often perceived as such. The last four are actual failures that are due to transactions costs and limitations on contracting possibilities.

Type 1: Retrospective Disappointment

The entire economy goes down, so that even with perfect contingent claims markets everyone gets less. People may feel themselves to be involuntarily unemployed if, as in the Great Depression, their best wage opportunities come from selling apples. The question they ask themselves to determine whether they are "involuntarily unemployed" is not whether

they would work at the present going wage, but whether they would have done so at the previous going wage.[33]

Type II: Position Reversal

There is perfect labor contracting, but because there are changes in comparative advantage over the business cycle and because the appropriate fallback opportunity for some individuals is an activity that is recorded as unemployment (e.g., own home repair), there is variability in employment levels. With a heterogeneous labor force, some employees will be dismissed while others remain working. The dismissed workers, feeling that they have been treated differently from retained workers, are likely to consider their unemployment involuntary. They may make the natural mistake of asking whether they would work at the wage now being received by individuals formerly earning no more than they, not recognizing that the changed state of the world has changed their relative earnings opportunities.[34] Cherished rank orderings may be reversed (e.g., university professors may earn less than plumbers when both drains and faculty positions fail to flow). The less risk-averse are more likely to find themselves in lowered positions during poor times (and may complain less, too).

Type III: Interstitial Unemployment

Firms are heterogeneous; therefore, efficiency requires that workers switch from firm to firm. Thus, despite otherwise perfect labor contracting

there is some transitional unemployment which would not exist if contingent claims markets worked perfectly and without cost.

Type IV: Underemployment Due to Second-best Contractions

Severance pay guarantees best behavior by the firm. Moreover, to tie themselves to the firm in good times so as to provide risk spreading, workers leave some "earnest money" with the employer. This surety bond, which may take the form of seniority privileges, nonportable retirement benefits, or a wage stream rising faster than productivity, represents a second-best situation for the worker, because it imposes a high cost on transition between jobs.

Given this limitation, it will not be worthwhile for the worker to commit himself as fully as he would with an optimal contract. Consider a situation just sufficiently prosperous that the worker leaves although he would have stayed under optimal contract. Although his departure will lead him to a higher wage, it is likely that he is moving to a position where his value added is lower. Thus, there is a deadweight loss in productivity. This strict market failure leads to involuntary underemployment, though the worker may not recognize it as such, since he is actually moving to a higher wage position. The worker has paid for the deadweight loss through lower wages while he was at the first firm.

Type V: Lump-Sum Severance Pay

There is perfect labor contracting. Severance pay is used to spread risks and to ensure that the firm does not try to violate its

contractual obligations. But severance pay is made in a lump sum, as opposed to a period-by-period stream of compensation that continues until the individual is again employed at this old wage level. Such a lump-sum arrangement offers two advantages. The firm need not monitor a dismissed worker's activities, and a worker is not discouraged from taking another job.

The lump-sum approach also has disadvantages, both real and psychological. The real disadvantage is a loss of risk spreading. The lump sum is a prize equal to an expected string of small losses representing the difference per period between previous salary and salary after layoff. This arrangement puts the worker at risk. He wins in the lottery if he is below his old wage level for a shorter than expected period of time, and loses if it takes him longer to recover his income.

The psychological disadvantage arises because the lump-sum payment becomes a sunk gain as soon as it is received. Workers therefore are unhappy about not being employed at their old positions, for on a period-by-period basis their earnings will be lower. Workers who prefer not to remain at their fallback opportunities may report themselves as being involuntarily unemployed, even though their present employment status is what they contracted for and should have contracted for.

Type VI: Layoffs Due to Limitations on Contracts

There are only limited opportunities to contract using severance pay, rising wages, and seniority privileges. For any of a variety of

reasons (minimum wage, fair play, union contracts, workers' desire to defend wage differentials), wages cannot be cut. When times become relatively bad, the firm just dismisses workers, even though with appropriate severance pay it would not.

When there is imperfect information (or imperfect contract enforcement), agents realize that agreements will hold only if they are both ex ante and ex post rational. Types IV, V, and VI of involuntary unemployment arise from the nature of the second-best solution.

CONCLUSION

We have enriched the standard formulations of labor contracting between employees and employers by taking explicit account of heterogeneity of workers and firms. Workers differ in their preferences and in their productivities at different firms in different states of nature. Firms differ in the way they are affected by the business cycle, and in their ability to make employment commitments to workers. Given this heterogeneity, particularly that among firms, efficiency requires that workers switch among firms as business conditions change. Risk aversion on the part of workers, however, may require them to stay with firms over protracted periods of time, or over a wide range of business conditions, thereby providing a means to average their wages.

Unfortunately, there may be impediments to contract on the part of both firms (the information on which they base employments and layoffs may not be objective and verifiable) and workers (they cannot indenture

themselves). Apart from whatever other roles they may play, the well-known labor market institutions of severance pay, pension and retirement benefits, and a lifetime income curve that rises more swiftly than productivity help to overcome limitations on contract. In particular, they make fixed-wage labor contracts ex post rational. These institutional arrangements are consistent with rational maximizing behavior in a world where the ability to make and keep commitments is limited.

Our formulation highlights the role of comparative advantage and the possibilities for and limitations on contracting in providing a variety of explanations for (i.e., types of) involuntary employment. Some types are the product of market imperfections; others merely reflect understandable worker attitudes. The next step is an empirical analysis of our ten testable hypotheses, to gauge the relevance of our model. To the extent it does explain real-world phenomena, the policy implications for such areas as stabilization policy, unemployment insurance, manpower training programs, and pension policy will be significant.

NOTES

[1]For an excellent survey of the implicit contract literature, see Azariadis (1981) and Akerlof and Miyazaki (1980). Among the more important articles in the literature are those by Baily (1974), Azariadis (1975), Mortensen (1978), Grossman (1978), Hall and Lilien (1979), Lazear (1979), Chari (1980), Grossman and Hart (1981), Solow and McDonald (forthcoming), Bean (1981), and Green (1980).

[2]See Solow (1980) for a charming distillation of these explanations. See also Varian (1976) and Malinvaud (1977).

[3]If this extreme assumption were relaxed, our results would change quantitatively, but not qualitatively. Even if savings were possible, firms, which confront interest rates well above those paid on individual savings accounts, should do considerable saving on behalf of workers. Azariadis (1981) discusses this issue.

[4]We assume that the firm provides unemployment insurance. If it is provided by the government, there must be accurate experience-rating of firms; that is, they must be charged on the basis of their record. Otherwise, there will be a severe moral hazard problem: firms will abuse the system and lay off workers too readily.

[5]These institutions may promote wage rigidities in other ways. For example, unions may see themselves as representing their own long-term members more than workers as a whole; if so, given seniority, they will prefer employment cuts to wage cuts. The rules of the collective bargaining game usually allow labor to set wages and firms to get employment levels.

The contracted wage may not shift much during a recession; it will not shift at all if demand shrinks by the same amount at every wage level (Solow and McDonald, forthcoming).

[6]Depending upon one's methodological preferences, it may not be important for individuals to <u>believe</u> in an implicit contract theory provided that they <u>behave</u> as if they did (Friedman, 1953). Because there are many theoretical explanations of involuntary unemployment, it is important to develop a series of empirically testable hypotheses to gauge the strength of each approach's predictive power.

[7]The widely discussed labor practice, common in Japan, of lifetime employment contracts without layoffs would be optimal if the worker's productivity in the poorest states were sufficient to offset his disutility of work, and if across states of the world this productivity would always be highest at the same firm.

[8]Workers have information about themselves that cannot be shared with the firm without costs (e.g., transition costs, risk aversion, and fallback opportunities). A variety of firms, each offering its own specialized contract, is a good substitute for each firm providing individualized contracts. Workers--at least early in their careers--may choose between very cyclical employment offering high wages, as in the automobile industry, and jobs with lower wages and greater security, as in civil service positions.

[9]If this assumption were not satisfied, it would be quite reasonable for an individual to wish to have a higher income attached to employment situations where his value-added was lower. For example, the small-time

building contractor who becomes a tennis pro when the construction business is slack might prefer to have his contractor's income coupled with his athletic employment and vice versa, given the greater free time he has as a tennis professional. Because of the types of enforceability problems we shall be considering below, few employment contracts offer an individual greater income as an accompaniment to greater leisure or less onerous employment.

[10] In certain second-best situations, because of the absence of perfect contracting capabilities, the worker might actually move from one identical firm to another over different states of the world. One firm might hold him for very good states, another for medium, and another for bad.

[11] Big firms may be able to switch workers from job to job or even from division to division. Such switching allows for contracting over a broader range of states, hence more risk spreading for the workers. This provides an economy of scale to the firm, and in particular offers an advantage to diversification, as for example with a conglomerate. Even if asset markets were perfect, so that stockholders were in effect risk neutral, firms would have an incentive to diversify to protect their workers.

[12] Public information, i.e., information that both workers and firms know they can monitor in common, could sometimes prevent such deception. For example, if an industry is doing well, a firm within it might have a hard time convincing a worker that it was doing so poorly that he should seek employment elsewhere.

The limited enforceability problem might be ameliorated by having a worker's wage at a firm vary depending on public information. Additional risk due to wage variability would be accepted in trade for reduced risk due to limited contracting opportunities.

[13]This problem has been effectively addressed by Grossman and Hart (1981) in the context of a risk-averse firm. Workers demand a fixed wage and let the firm choose how many workers to employ. The only way in which workers can share some of the risk is to permit their employment to be correlated with the state of nature. Thus they receive too high a wage when they are employed and too low a wage when they are unemployed.

[14]See Azariadis (1976), Bean (1981), and Barro (1977) for a greater discussion on this point of ex post rational contracts and moral hazard. It is sometimes alleged that firm reputations can alleviate the moral hazard problem. Workers can observe the experience of numerous others who have preceded them in employment. The reputation effect would partly explain why large firms have a smaller proportion of layoffs than small firms.

[15]Pratt and Zeckhauser (1981) show that despite asymmetric information, it will generably be possible to induce agents to report information honestly and/or take appropriate actions so as to produce a first-best outcome if (a) there is a transferable medium of exchange, (b) agent i's private information does not directly enter agent j's utility function, and (c) agents are risk-neutral.

Requirement (a) is satisfied by the availability of severance pay. Requirement (b) is met since no firm's well-being is an argument of any worker's utility function. Requirement (c) is met for firms, who can thus be induced to inform and act appropriately, but it is not met for workers. Interestingly, this turns out not to be a problem. Optimal severance pay will eliminate all risk for workers, so risk aversion induces no losses.

Risk aversion could defeat a first-best, fully cooperative outcome--as we shall soon see--if workers must commit themselves not to leave a firm in some circumstances where they could earn more elsewhere. Even then, if transfers could be made on the basis of public information, i.e., what the worker and firm both know, a first-best full risk-spreading outcome is achievable.

[16]This result no longer holds if the worker's marginal utility of income depends upon his employer or the state of the economy as it will not be optimal to equalize his income across states. (See footnote 9.)

In multiperiod models, paying severance pay in a lump sum leaves residual risk, because the worker's time until he is recalled is uncertain, as is his loss of wages.

[17]See Medoff and Abraham (1981) who conclude: "The new evidence presented in this study strongly supports the claim that seniority independent of productivity plays a major role in the compensation and termination decisions affecting all employee groups in most U.S. workplaces." Ioannides and Pissarides (1980) show that a monopsonist can stop other firms from stealing its employees by instituting a rising wage path. Workers who have a vested interest with a firm are less likely to be lured away by competitors. A worker who banks a stockpile of "savings"

with a firm--i.e., accumulated differences to date between productivity and wages--takes the risk of being fired. This suggests that earnings gradients will be steepest at firms that establish reputations as secure employers.

[18]Efficiency requires this departure if value added is sufficiently higher. Even in one profession that permits indentured servitude, the military, financial penalties allow for some shifting of resources. The Air Force, for example, allows its pilots to leave if they pay a training cost reimbursement.

[19]See Appendix 1 for a discussion of a two-period model of worker commitment. We shall not discuss here other advantages that may flow from worker commitment. For example, it may lead the firm to provide him with more training in skills that are transferable and may make the firm more willing to trust him with confidential information.

[20]Arnott, Hosios, and Stiglitz (1980) discuss the tension between labor mobility and employment insurance when wages are no longer a perfect signal of relative productivity because of the reduced variability embodied in the implicit contract.

[21]The value of these benefits would reflect on an actuarial basis the likelihood of leaving. An individual who had set aside $1,000 with a one-half chance of leaving would receive $2,000 in benefits should he remain. Such an actuarial adjustment would be essential if optimal contracting would require the worker to leave in some states of the world when his comparative advantage elsewhere was sufficiently high.

[22]The example of professional baseball is instructive. The field is extremely risky, yet players do not find it optimal to put aside a sufficient [illegible]ety bond to keep them attached to their team and thereby help smooth earnings, and they recently pushed hard (and successfully) to eliminate legal ties due to the reserve clause. (This also had the advantage of enabling them to break a contract under which they had received substantial advance money, a ploy based upon the folly of our legal system and unrelated to the theory of involuntary unemployment.) Aside from the disincentive effects on insuring earnings, the explanation for the absence of tying would seem to be that a player's relative contribution to different teams is likely to change rapidly. Reallocation of a player across teams is important. The net result is that professional teams have to pay their high performance players extraordinary salaries in their top years to keep them from being bid away. Contracts tend to offer small severance pay, large salaries in good years, and limited incentives to stick with the firm after the required six-year period.

[23]Heterogeneity in quit rates could also explain this phenomenon. Even if all workers had constant drop-out and drop-in rates across time, long-term as opposed to short-term participants in the labor force would be disproportionately individuals with low drop-out rates. (Spilerman, 1972 develops a mover-stayer labor force participation model with heterogeneous probabilities.)

[24]Since unionized firms pay more than nonunionized ones, their workers are less likely to leave. Other unionized employment is unlikely to be

available, because the higher wages will create a queue. As David Ellwood pointed out to us, since unionized workers are less tempted to look elsewhere, they will have less need for a steep earnings gradient. This provides one counterbalance to the incentive for long-term union workers to exploit younger workers by negotiating a steep age-earnings profile.

[25]An intriguing complementary explanation is that wages starting below and rising above value added may in effect represent workers' investment of their "savings" with their employer. Because of tax distortions and economies of scale in investment, firms will have better investment opportunities than do workers.

[26]Ordering employment assignments at firms over states by the critical ratio method, that is, by comparative advantage W_{i1}/W_{i2}, assures the worker that he reaches a constrained optimum of the two (n) expected total wages at the two (n) firms (Weinstein and Zeckhauser, 1973). With a constant fallback wage, though not in general, this will correspond with ranking by absolute advantage, $W_{i1} - W_{i2}$.

With risk aversion, it may not be optimal to reach a constrained optimum of expected total wages. The utility-maximizing allocation of work effort may be strictly inferior to another that offers a higher expectation of earnings at all firms. The two factors that constitute the expectation--probability of being at the firm and wage--matter, not merely their product.

[27]There is also reason to believe that high-paying jobs go to workers with specialized skills that are inelastically supplied (Rosen, 1981).

[28]Some universities have both tenured and non-tenured full professorships. Our hypothesis would be that the more versatile new appointees would opt for the higher pay and lower security of the nontenured professorships. Versatility probably, but not necessarily, correlates with absolute capability.

[29]This result is modified if we introduce the mathematical trick of probabilistic contracting. In state 1, Holly would contract to be with Firm A X percent of the time and with Firm B (1-X) percent of the time. In state 2, Holly would work for Firm A Y percent of the time and would spend the other (1-Y) percent of her time with Firm C. Optimal portfolio theory suggests that an individual will always "purchase" some positive amount of employment with the firm at which he has the highest productivity in each state. Probabilistic contracting is theoretically appealing but of little practical importance.

[30]Such a well-ordered shuffle would reflect a considerable degree of regularity in employment opportunities. Regularity might be observed, for example, if firms varied in their amounts of capital per worker, so that the highest-quality workers were hired by the most capital-intensive firms, and if a change in market conditions just shifted the total number of workers (adjusted for quality) a firm would wish to hire. If there were not such regularity, worker quality would be ambiguous, and although worker A might earn more than B in good times, in bad times the ordering might be reversed.

[31]Proposition 2 1/2, limiting property taxes in Massachusetts, provided an interesting application. Conditions for all public servants

were dramatically reduced at one time. It was therefore not surprising that in a number of communities workers accepted reductions in wages in preference to reductions in the number employed.

[32]Workers will get paid more initially if they are likely to forfeit benefits in this way. The efficiency loss derives from the creation of risk and the destruction of arrangements that tie workers to firms, the so-called "Golden Handcuffs" ([London] Sunday Times, 1981).

[33]Observable declines in physical productivity may be viewed differently from declines in aggregate demand. Farmers in an agricultural nation that is hit by a drought, although forced to turn to alternative lower-paying employment or struggle by with their traditional crops, may not feel themselves to be involuntarily unemployed.

[34]Their perceptions may be more gloomy still if they now switch to a firm where their period wage is lower because they are then at the top of their earning opportunities with that firm. Consider an individual who is worth $9 at firm A during very good times, $7 at firm A during good times, $5 at firm B during moderate times, and $3 at firm B during bad times. With implicit contracts, assuming the four states are equally likely, he will receive $8 at firm A and $4 at firm B. The drop from good to moderate times--though his productivity is down by less than 30 percent (from 7 to 5)--cuts his wage in half.

See the example in Appendix 1, where a worker switches to a job where he is more productive, yet receives a lower wage.

APPENDIX 1: THE OPTIMAL CHOICE OF EMPLOYMENT CONTRACT

This appendix further illustrates the agent's optimal choice of an employment contract. Holmstrom's (1981) elegant example with a single firm is followed by the solution to the more difficult problem when there are two firms. For simplicity, firms are assumed able to engage in perfect contracting, but workers must provide earnest money that forms the basis of their insurance contract.

The states of nature, s, are ordered by the size of output. In state s, the value of a worker's production at each of the many identical firms is $W_{ij}(s) = s$. In the simple example of a two-period model, a worker is only able to obtain insurance in the second period. This follows from the fact that were a firm to guarantee a salary of $\bar{y}$, no worker would ever go to that firm if s were greater than $\bar{y}$. No firm can afford to pay any positive $\bar{y}$ only to workers who are worth less than $\bar{y}$.

In the first period a worker is offered a salary equal to the value of his output, s_1. He then has the option of leaving a certain amount of his salary, z, with the firm to provide the funding for an insurance contract. In the second period, the firm must still pay the worker at least the value of his output. The insurance money, z, is paid back to the worker on an actuarially fair basis to boost his earning in bad states of nature.

In any state in the second period in which the worker would earn less than in the first, the worker would like to transfer income from

the first to the second period where he has a higher marginal utility of income. He will continue providing funds for second-period consumption until there are no states in which he would be consuming less in the second period than in the first. See Figure A.1.

There is no mechanism for averaging income in the states in which the worker is worth more than $s_1 - z$, and thus in those states he earns the value of his output, s_2. The amount of insurance demanded is a monotonically increasing function of the first period salary ($\partial z/\partial s_1 > 0$). Provided $U'' < 0$, these arguments do not depend on the extent to which the agent is risk averse.

When there is more than one firm, a worker must make a tradeoff between production efficiency and insurance. This point is illustrated by reference to the two production possibilities, $W_{i1}(s)$ and $W_{i2}(s)$, depicted in Figure A.2.

The worker is initially employed by firm one and is offered his competitive value, $W_{i1}(s_1)$. If his first period earnings are small, then he will not seek very much insurance and the story is essentially as before; see Figure A.3. Firm 1 will average his salary in second period until he is worth more than $s_1 - z$. From then on he will just be paid the value of his product. He will switch to firm 2 when his [illegible]put there is greater ($s > X$). The optimal $C(s_2)$ is denoted by the [illegible]y line.

When the worker earns a sufficient salary in the first period to seek insurance with the first firm past the efficient switching point,

First period consumption $s_1 - z$

Second period consumption $C(s_2)$ $\begin{cases} s_1 + z & s_2 \leq s_1 - z \\ s_2 & s_2 \geq s_1 - z \end{cases}$

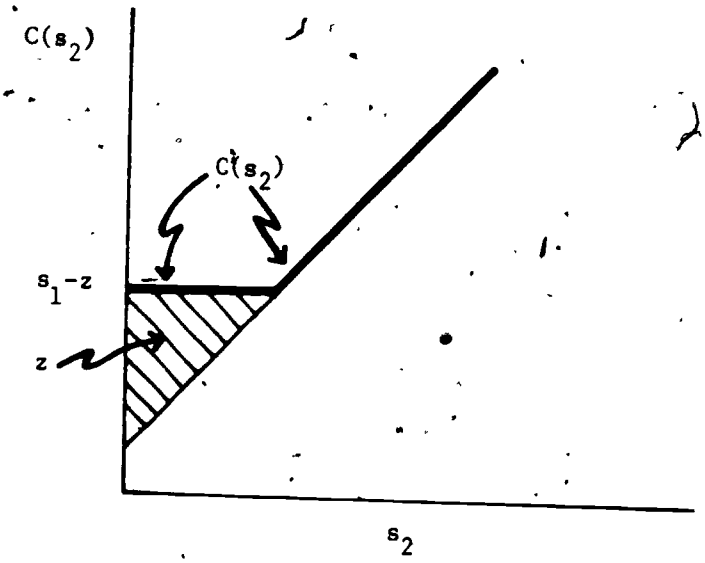


Figure A.1

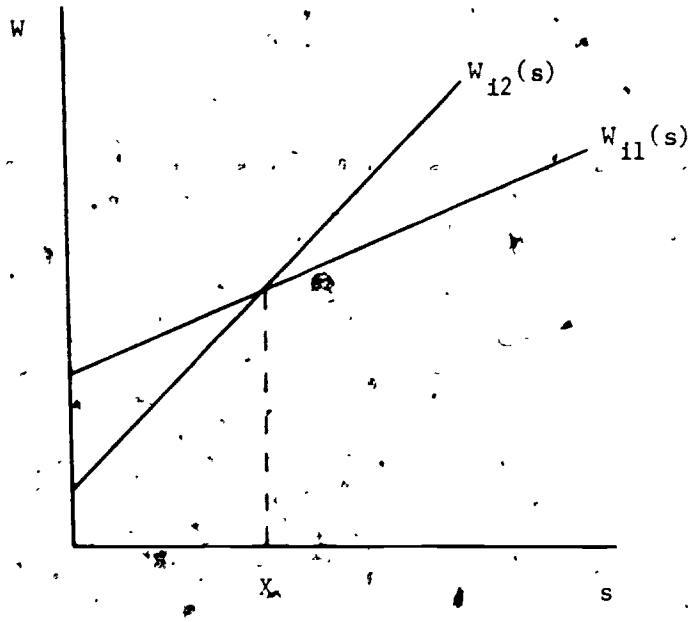


Figure A.2

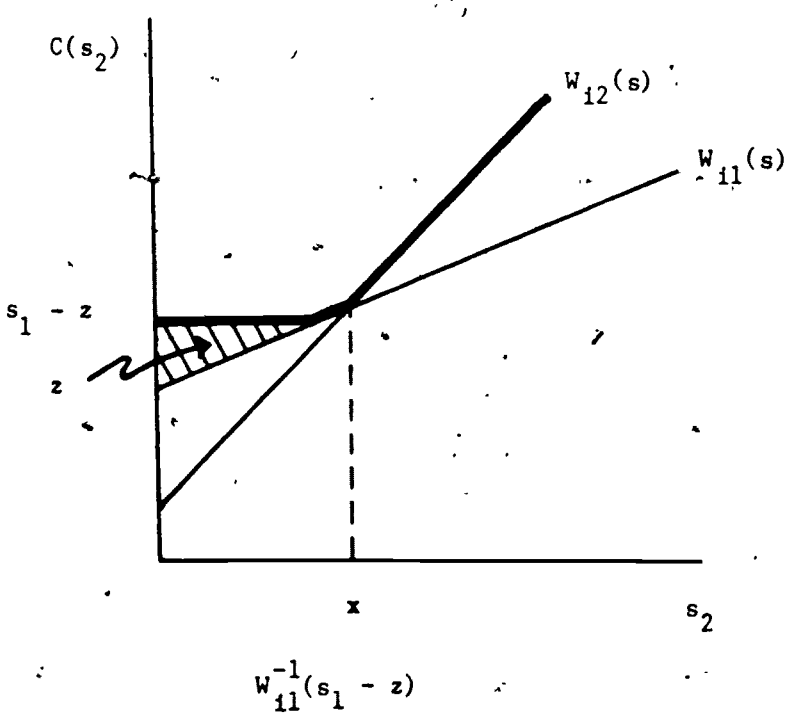


Figure A.3

X, the story is more complicated; see Figure A.4. The optimal $C(s_2)$ is denoted by the heavy line; on it X* represents the state up to which it is better to work at firm 1 than firm 2 in period 2.

Just as in the single-firm story, as long as the worker stays with the first firm, he will always earn a salary of at least $s_1 - z$. This follows from the fact that in order for the worker to be kept at firm 1 in state s he must be paid at least max $[W_{i1}(s), W_{i2}(s)]$. If this salary is less than his first period income, then it is desirable to transfer income into the second period on an actuarially fair basis.

There is, of course, a difficulty with transferring income into the second period past X. To do this, a worker must sacrifice potential output of $W_{i2}(s) - W_{i1}(s)$. At the optimum switching point, this loss in efficiency will be exactly balanced by the gain from income averaging, since the marginal utility is higher in the second period. The first-order conditions determining the optimal z to transfer into the second period, and the optimal switching point X* are given by

$$\text{i)} \quad U(s_1 - z) - U(W_{i2}(X^*)) = U'(s_1 - z)[(s_1 - z) - W_{i1}(X^*)], \text{ and}$$

$$\text{II)} \quad \int_0^{X^*} [s_1 - z - W_{i1}(s)]p(s)ds = z.$$

It is clear that a worker when switching from firm 1 to firm 2 at X^* will feel injured, perhaps even involuntarily unemployed; he is being forced to accept a pay cut even though he is transferring to a job where his labor is more productive.

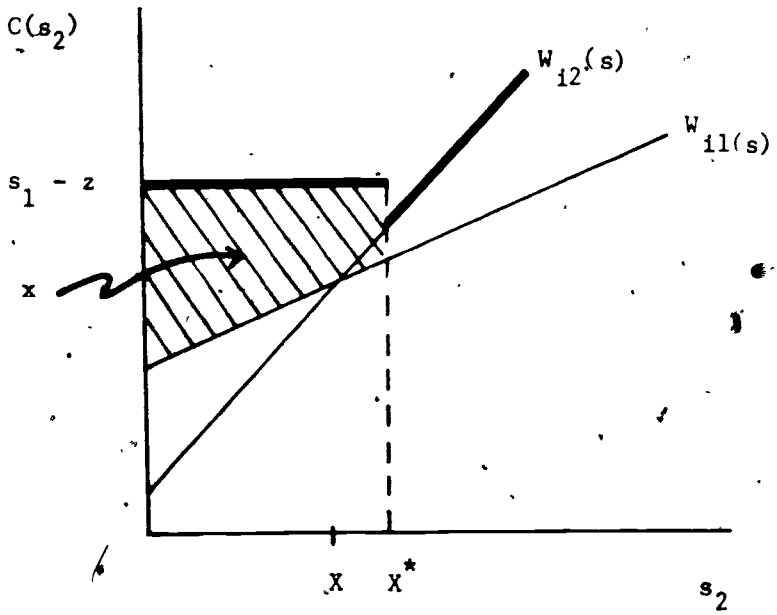


Figure A.4

APPENDIX 2: CONTINUOUS BLOCKS OF EMPLOYMENT

A worker employed at firm 1 chooses a minimum value added, $\underline{s}$, at which to switch to firm 2. Once with firm 2, he must also decide upon a minimum value below which he should switch back to firm 1. This exit state, x(s), will depend on his value when he entered firm 2. Let (2,s) denote working for firm 2 having entered state s.

From \ To	Firm 1	Firm (2,s)
1	$\underline{s}/20$	$1/20$
(2,s)	$x(s)/20$	$1-x(s)/20$

Define

P_1 as steady-state frequency of employment with firm 1,

$P_2(s)$ as density function for steady-state frequency of employment with firm 2 having entered in state s, and

$C(s)$ as salary at firm 2 with entry at s.

Then

$$C(s) = \frac{x(s)}{20} \cdot s + [I - \frac{x(s)}{20}] [\frac{20 + x(s)}{2}],$$

$$P_2(s) = p_1/x(s), \text{ and}$$

$$P_1[1 + \int_{\underline{s}}^{20} 1/x(w)\, dw] = 1.$$

The worker's average utility (no discounting) is based on the steady-state frequencies from the Markov process described above. The worker chooses $\underline{s}$, and an $x(s)$ schedule to maximize

$$EU = P_1 U(12) + \int_{\underline{s}}^{20} P_2(w)\, U[C(w)]dw.$$

The first-order conditions satisfy:

$$x(s): \quad EU - U[C(s)] + x(s)U'[(C(s)][s - x]/20 = 0, \qquad (2.1)$$

$$\underline{s}: \quad EU - U[C(\underline{s})] = 0. \qquad (2.2)$$

Combining (2.1) and (2.2), we see that at the lowest entry level, $\underline{s}$, the exit level is the same as the entry point, $x(\underline{s}) = \underline{s}$.

Further differentiation of (2.1) with respect to s shows that for any risk-averse utility function [$U'' < 0$],

$$x'(s) = \frac{x(s)[s - x(s)]U''[C(s)]}{20U'[C(s)] - U''[C(s)][s - x(s)]^2} < 0.$$

Thus the range of employment states is greater when the worker enters in a relatively prosperous period. Intuitively, the greater the surplus, s, to be distributed, the higher is the salary while with the firm and the broader is the range of states over which the surplus is spread.

REFERENCES

Akerlof, G. and H. Miyazaki. 1980. "The Implicit Contract Theory of Unemployment Meets the Wage Bill Argument." Rev. Econ. Stud., (January), 321-338.

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

Azariadis, C. 1975. "Implicit Contract and Unemployment Equilibria," J. Polit. Econ., (December), 1183-1202.

_____. 1976. "Asymmetric Wage Behavior." In O. Kyn and W. Schrettl (eds.), On the Stability of Contemporary Economic Systems. Gottingen: Vandenhoeck & Ruprecht.

_____. 1981. "Implicit Contracts and Related Topics: A Survey." In Z. Hornstein et al. (eds.), The Economics of the Labour Market. London: HMSO.

Baily, M. N. 1974. "Wages and Employment under Uncertain Demand." Rev. Econ. Stud., 41 (January), 37-50.

Barro, R. J. 1977. "On Long-term Contracting, Sticky Prices, and Monetary Policy." J. Mon. Econ., July.

Bean, C. 1981. "Optimal Wage-Bargains." M.I.T., Ph.D. Dissertation, March.

Chari, V. V. 1980. "Involuntary Unemployment and Implicit Contracts." Northwestern University, The Center for Mathematical Studies in Economics and Management Science, Discussion Paper 459, August.

Friedman, Milton. 1953. "The Methodology of Positive Economics." In Essays in Positive Economics. Chicago: University of Chicago Press.

Gordon, R. 1981. "The Consumer Price Index: Measuring Inflation and Causing It." The Public Interest, 63 (Spring), 112-134.

Green, J. 1980. "Wage-employment Contracts." University of Pennsylvania, Center for Analytic Research in Economics and the Social Sciences, Working Paper, September.

Grossman, H. 1978. "Risk Shifting, Layoffs, and Seniority." J. Mon. Econ., 4 (November), 661-686.

Grossman, S. and O. Hart. 1981. "Models of Labor Market Equilibrium." Amer. Econ. Rev., 71 (May), 301-307.

Hall, R. E. 1981. "The Importance of Lifetime Jobs in the U.S. Economy." National Bureau of Economic Research, Working Paper.

Hall, R. E. and D. M. Lilien. 1979. "Efficient Wage Bargains under Uncertain Supply and Demand." Amer. Econ. Rev., 69 (December), 868-879.

Holmstrom, B. 1981. "Contractual Models of the Labor Market." Amer. Econ. Rev., 71 (May), 308-313.

Ioannides, Y. and C. Pissarides. 1980. "Wages and Employment with Firm-specific Seniority." Boston University, mimeo, October.

Keynes, J. M. 1936. The General Theory of Employment, Interest, and Money. New York: Harcourt, Brace.

Lazear, E. 1979. "Why Is There Mandatory Retirement?" J. Polit. Econ., 87 (December), 1261-1285.

Lazear, E. 1980. "Agency, Earnings Profiles, Productivity and Hours Restrictions." University of Chicago, mimeo, December.

Malinvaud, E. 1977. The Theory of Unemployment Reconsidered. Oxford: Blackwell.

Medoff, J. L. and K. G. Abraham. 1981. "The Role of Seniority at U.S. Work Places: A Report on Some New Evidence." Harvard Institute for Economic Research, Discussion Paper no. 809, January.

Mortensen, D. 1978. "On the Theory of Layoffs." Northwestern University, Discussion Paper no. 322.

Pratt, J. and R. Zeckhauser, 1981. "Expected Externality Payments: Incentives for Efficient Decentralization." Harvard University, mimeo.

Rosen, S. 1981. "The Economics of Superstars." University of Chicago, mimeo.

Solow, R. 1980. "On Theories of Unemployment." Amer. Econ. Rev., 70 (March), 1-11.

Solow, R. and I. McDonald. Forthcoming. "Wage Bargaining and Employment." Amer. Econ. Rev.

Spilerman, S. 1972. "Extensions of the Mover-stayer Model." Am. J. Sociol., 78, 599-627.

The [London] Sunday Times, 1981. "Pensions: Scandals of the Golden Handcuffs." Part I, May 17, p. 63, Part II, May 24, p. 61.

Varian, H. 1976. "Keynesian Models of Unemployment." M.I.T. Dept. of Economics, Working Paper.

Weinstein, M. and R. Zeckhauser. 1973. "Critical Ratios and Efficient Allocation." J. Pub. Econ., 147-158.