FLUCTUATIONS IN EQUILIBRIUM UNEMPLOYMENT

Robert E. Hall
NBER and MIT

Working Paper No. 226

NATIONAL BUREAU OF ECONOMIC RESEARCH
1737 Cambridge Street
Cambridge, Massachusetts 02138

December 1977

This research was sponsored by the National Science Foundation.

Most recent thought about unemployment views the unemployment rate as fluctuating around a constant equilibrium or natural rate. Unemployment rises above the equilibrium rate when demand is unexpectedly low and falls below when it is unexpectedly high. Fluctuations in unemployment are interpreted as symptoms of disequilibrium--as participants in the market become aware of conditions in the market and as they are able to adjust the terms of employment arrangements, wage movements offset unexpected movements of demand and equilibrium is restored. The continuous pattern of fluctuations of unemployment reflects the continuous arrival of unexpected shocks in demand. The major exceptions to this view are the universally recognized shift in equilibrium unemployment attributable to the changing composition of the labor force and the more controversial claim that various government programs including unemployment insurance have raised the equilibrium unemployment rate in recent years.

The inspiration for this paper is a body of evidence that points rather strongly in the direction of a larger role for fluctuations in equilibrium unemployment than is generally recognized. Efforts to partition historical movements in unemployment into equilibrium and disequilibrium terms have, in some cases, attributed the great bulk of movements to the equilibrium term, though this finding has not been emphasized or attracted much attention. Similarly, movements of wages are so weakly associated with levels of unemployment that the interpretation involving equilibrating fluctuations in wages requires an elasticity of demand for labor that is absurdly high. Finally, the

pattern of unemployment rates in specific geographical and occupational labor markets seems to require an equilibrium interpretation of an important part of their movements relative to one another, unless disequilibrium can persist for many years.

Fluctuations in the equilibrium rate of unemployment can only be understood within a theory of the natural or equilibrium rate. It is not enough to say that unemployment is the difference between supply and demand in the labor market, though of course it always will be. In equilibrium, no participants in the market can have an unexploited opportunity to make themselves better off. At the equilibrium unemployment rate, employers cannot obtain labor at lower cost by offering work at below the market wage to the unemployed. Unemployed workers cannot raise their effective real incomes by taking lower wages in exchange for immediate employment. The task of the theory is to explain why any unemployment remains at all when these conditions are satisfied. Part of this problem has been studied in detail in the "search theory" of unemployment--once a worker becomes unemployed, it is reasonably well understood why the worker does not become employed again immediately. The theory of why people become unemployed in the first place is less well developed and is the main concern of this paper. Most of the unemployed are looking for new work because their previous jobs ran out. Consequently, the main ingredient of a theory of the flow of workers into unemployment is a theory of the duration of employment. Such a theory is developed here, along reasonably standard lines.

Within the theory, both employers and workers care about the duration of employment. Duration can be viewed as a characteristic of a job along with its wage. Then an efficient employment contract sets a duration and a wage at a point where the isocost curve is tangent to the indifference curve. Such a point is a desirable compromise between the employer's desire to retain flexibility over future levels of employment and the worker's interest in stable employment.

The paper unites this theory of the flow into unemployment with a simple model of unemployment. The resulting model of the labor market does indeed have an equilibrium where the unemployment rate is positive--under the efficient employment contract, jobs have finite lengths and workers are continually moving through the labor market. But the equilibrium is indeterminate. The market may be in equilibrium with slack conditions and high unemployment or tight conditions and low unemployment. There is a socially optimal equilibrium, generally with very tight conditions, but the self-interests of participants in the market by themselves will not push the market toward the optimum. The basic difficulty is that a single small employer is incapable of assuring a prospective employee of a favorable job-finding experience after the job runs out. However there is a strict upper bound to the equilibrium unemployment rate: If conditions in the open labor market are bad enough, employers will respond by offering permanent jobs. When all employers do this, unemployment vanishes. Severe depressions cannot be equilibria in terms of the model.

Indeterminate unemployment rates fit well with the evidence of large unexplained differences in unemployment across cities and unexplained

shifts in relative unemployment rates over time. Of course, there is an alternative explanation of these findings, which lies behind almost all popular accounts and strongly influences economists as well--labor markets are in perpetual disequilibrium, observed differentials reflect alternative levels of demand, and wages never adjust to clear the markets. Under this explanation, employers are ignorant of the possibility of obtaining labor at below the prevailing wage in slack markets or they are prohibited from doing so. The believer in permanent disequilibrium and permanent unexploited opportunities for profit will not be convinced by this paper that there are any mysteries about unemployment. Rather, the paper does offer a possible alternative explanation of the facts that rests on economic equilibrium and invokes no failure of the principle that individuals follow their own self interests.

Aggregate evidence about the relative importance of equilibrium and disequilibrium fluctuations in unemployment

A partitioning of the variance of measured unemployment into equilibrium and disequilibrium terms is possible if sufficiently strong assumptions are made. There do not seem to be any studies that have focused directly on this question, but highly relevant evidence is presented by Sargent (6,7) and Barro (1). The essential idea in both cases is to identify the disequilibrium term with the unexpected movement in some variable that is thought to be a good indicator of the excess supply of labor (prices in Sargent's work and the money supply in Barro's). The residual plus any other term in the equation then measure the equilibrium unemployment rate. The results of this kind of analysis give an unambiguous partitioning if the disequilibrium term has only a contemporaneous effect and no lagged effect. Sargent's equation 1 in (7, p. 235) is a good place to start:

$$u_t = \underset{(.144)}{-.287\, x_t} + \underset{(.0017)}{.0043} + \underset{(.0000014)}{.0000007\, t} + \underset{(.11)}{1.47\, u_{t-1}}$$

$$- \underset{(.20)}{.59\, u_{t-2}} - \underset{(.30)}{.03\, u_{t-3}} + \underset{(.13)}{.04\, u_{t-4}} \; ;$$

$$R^2 : .908 \qquad \hat{\sigma} : .371$$

u_t is the aggregate unemployment rate and x_t is a measure of unexpected price inflation, constructed as the log of the price level predicted by contemporaneous exogenous variables and lagged endogenous and exogenous variables less the log of the price level predicted using only lagged variables. The variable x_t measures all of the new information contained in the most recent exogenous variables that is relevant for prices. By construction, it is uncorrelated with any of the other variables in the equation. Consequently, the variance in u_t can be decomposed into three unambiguous components: one associated with x_t, interpreted as the disequilibrium component, one associated with all of the other independent variables, interpreted as the predictable part of the equilibrium component, and the residual variance, which is also part of the equilibrium component. In terms of percentages of the total variance of unemployment, this decomposition is:

Equilibrium component	0.41 percent
Disequilibrium component	99.59 percent
predictable part	90.39 percent
residual	9.20 percent

The various predictors that jointly explain over 90 percent of the variance cannot be assigned individual contributions, because they are correlated with each other, but it is clear that the single best predictor is the lagged value of unemployment itself. Since this variable is known one quarter in advance, it cannot contribute to the disequilibrium term. Clearly, under the strong assumption that all disequilibrium influences have their effect only within the contemporaneous quarter,

essentially all of the fluctuations in unemployment are attributable to changes in equilibrium unemployment and none to disequilibrium.

By now almost all proponents of disequilibrium theories of unemployment believe that expectation errors have effects that last more than one quarter, either because information takes more than three months to diffuse through the labor market or because participants are bound by contracts which specify a quantity rather than a wage response to unexpected fluctuations in demand. The logic of this view suggests the inclusion of variables measuring errors in expectations for this quarter on the basis of information available two quarters ago, three quarters ago, and so on. Sargent does not seem to have estimated this kind of equation. In (6, p. 452), he reports a regression for the unemployment rate in which a large number of lagged endogenous variables appear, including prices, the money supply, government expenditures, and wages, as well as lagged unemployment. Since variables measuring expectation errors would be linear combinations of these variables, the regression includes all possible disequilibrium terms involving these variables. Still, lagged employment has almost as much explanatory power as in the simple equation reported above--u_{t-1} has a coefficient of 1.22 with a standard error of .14 and u_{t-2} has a coefficient of -.55 with a standard error of .21. This equation does not permit an unambiguous decomposition of the variance into equilibrium and disequilbrium components, but it is clear that the equilibrium movements captured by the lagged unemployment rates are an important part of the story. In a more recent theoretical paper, (8),

Sargent has argued that the persistent movements in unemployment set off by unexpected economic developments are precisely fluctuations in equilibrium unemployment.

Robert Barro's related work, (1), reaches a rather different conclusion. Using annual data, he finds that errors in expectations about monetary growth explain fluctuations in unemployment quite well. Barro uses the error made two years ago in predicting last year's monetary growth, not the error made two years ago in predicting this year's monetary growth, which fits in more closely with Sargent's model of an equilibrium process set off by a brief disequilibrium rather than the model of an extended disequilibrium because of contracts or slow diffusion of information. Barro also finds that the residuals from his unemployment equation are not at all serially correlated, so it is unlikely that adding lagged unemployment would much change the results. His equation does contain exogenous variables that make important contributions to fluctuations in the equilibrium unemployment rate--these are the fractions of men in the armed forces and a variable measuring the coverage of minimum wages. The equation leaves 22 percent of the variance in annual unemployment unexplained. Barro's evidence on the overall importance of shifts in equilibrium unemployment is ambiguous.

Another study, Hall (4), supports the view that fluctuations in the equilibrium unemployment rate are an important part of fluctuations in overall unemployment. That study proceeds by deriving a plausible distribution of contracts or information lags from the observed stochastic behavior of unemployment. This derivation is based on the assumptions

that the equilibrium unemployment rate is constant (after adjustment for demographic shifts) and that it is offsetting wage movements that restore equilibrium after a shock in demand. The model implies a Phillips curve relation between wage inflation and unemployment but is estimated without using any information about the covariation of the two variables. The slope of the implied Phillips curve is about four times as steep as the one found in the data. The paper concludes that forces other than equilibrating wage movements account for an important part of the fluctuations in unemployment. Movements of the equilibrium itself are the most likely candidates.

None of the aggregate evidence is at all conclusive. All of it amounts to saying that variables measuring disequilibrium have limited explanatory power in equations where unemployment is the dependent variable. The conclusion that all the rest of the movements in unemployment represent fluctuations in equilibrium unemployment rests on faith that the disequilibrium variables are doing their job. It is interesting that the money supply, which is frankly a measure of aggregate demand, is much more successful than Sargent's price variable, which he hypothesizes to be related to labor supply as originally proposed by Lucas and Rapping, and is also more successful than unexpected wage inflation, which is taken to be a measure of disequilibrium in the standard Phillips curve. One possibility which is completely consistent with all of the evidence is that unexpected monetary expansion can drive down the equilibrium unemployment rate.

Evidence from unemployment rates by city

An earlier paper of mine, (3), argued that much of the dispersion of unemployment rates across cities must reflect differences in equilibrium rates and not just temporary disequilibrium. The evidence is the extremely stable pattern of unemployment differences among cities over time. Chicago almost always has lower unemployment than does Los Angeles, and the reason does not seem to be that the demand for labor is stronger in Chicago. After a year or two, the adjustment of wages in Chicago or Los Angeles, the migration of labor, and the movements of employers would smooth out disequilibrium differences. But in fact the differences persist for much longer. My study, published in 1972, examined unemployment rates for 12 cities in 1966, when unemployment varied from 2.4 percent in Houston to 4.5 percent in Los Angeles. In 1974, eight years later, the pattern had remained very much the same, though unemployment was generally higher. The correlation of 1966 and 1974 unemployment rates was 0.69 across the 12 cities. The same 6 cities that were below the mean in 1966 were below in 1974 as well. Unemployment differentials across cities are extremely persistent--much more persistent than is the national unemployment rate. Only a small part of the differences are attributable to different industrial or occupational compositions of the labor forces of the cities, according to the earlier study. The data seem to require an equilibrium interpretation.

Evidence from occupational unemployment differentials

In any year, unemployment rates vary enormously by occupation--laborers, factory workers, and craft workers always have above average rates, and farmers and white collar workers have below average rates. This by itself is not any evidence that equilibrium unemployment rates can fluctuate; almost any theory will predict that different occupations will have permanent differences in turnover rates and so in unemployment rates. However, the structure of occupational unemployment rates tends to shift over time in a way that suggests rather strongly that the movements are changes in the equilibrium, not temporary disequilibrium shocks. Some of the shifts seem to be permanent--the best example is the sharp increase in the relative unemployment rates of professional and technical workers that occurred in 1970 and that has persisted ever since. Seven years ought to be long enough for wages to offset the changing demand and supply conditions and to restore balance in the market. The failure for the unemployment rate to return to the normal relation to the national unemployment rate suggests that the same forces that changed supply and demand also changed the equilibrium unemployment rate. Other shifts last for a number of years but are eventually reversed. For example, the unemployment rate among private household workers rose slowly relative to the national unemployment rate from 1959 to 1966 and then fell dramatically in 1970. Throughout this period, there was a steady decline of the labor force in household employment. Again, the slow rates of change seem incompatible with a disequilibrium process. In the household sector, the possible impediments to wage

adjustment through government intervention or labor unions are largely absent, so it is hard to see why wage rigidity alone could explain the behavior of unemployment in the sector.

It is possible to quantify the slow movements of occupational unemployment differentials in the following way: Let the permanent differences in occupational unemployment rates be measured by a set of coefficients, α_i , one for each occupation, i. Let the economy-wide influences, both equilibrium and disequilibrium, be measured by a set of coefficients, β_t , one for each year, t. Let the changes in the occupational unemployment structure be described by a residual, $\epsilon_{i,t}$. Then the serial correlation of the residual provides an indication of the rapidity of the changes. The coefficients α_i and β_t can be estimated from the regression,

$$\log u_{i,t} = \alpha_i + \beta_t + \epsilon_{i,t}$$

	Serial correlation of residuals (standard error)	Standard deviation of residuals, percentage points
Professional and technical workers	.85 (.13)	.14
Managers and administrators	.29 (.24)	.08
Sales workers	.24 (.24)	.05
Clerical workers	.94 (.09)	.07
Craft workers	.92 (.10)	.12
Operatives	.50 (.22)	.07
Non-farm laborers	.94 (.09)	.09
Private household workers	.70 (.18)	.12
Other service workers	.18 (.25)	.04
Farmers and farm laborers	.61 (.20)	.12
All occupations (pooled data)	.73 (.05)	.10

Four of the ten occupations have extremely high serial correlation coefficients, ranging from .85 to .94. For them, the hypothesis cannot be rejected that each random shift in unemployment relative to the national rate is a permanent shift (serial correlation of 1.0), in which case a

disequilibrium interpretation seems completely untenable. In three of the occupations, serial correlation is below 0.3 and is compatible with disequilibrium; however, these are also occupations where the residual after accounting for economy-wide effects is small. In the remaining three occupations, including the large and volatile category of operatives, the serial correlation of the departures from the national unemployment rate is comparable to the serial correlation of the national rate itself (about 0.6). Whether or not this level of serial correlation admits a disequilibrium interpretation is a matter of controversy.

For all occupations considered together, the serial correlation of shifts in unemployment relative to the fixed occupational structure and relative to economy-wide shifts is 0.73. In the year after a shift in the occupational unemployment rate occurs, almost 3/4 of the shift can be expected to remain; two years later, over one half can be expected, and so on. As a general rule, these shifts are sufficiently persistent to make exploration of the possibility of a changing equilibrium a useful undertaking, though of course a sufficient degree of wage rigidity throughout the economy could also explain the findings.

Job separations and unemployment

Though the popular conception of an unemployed worker is someone who has been at work until recently and is now looking for a new job, there are two other categories of unemployment that potentially deserve attention. First, as Martin Feldstein (2) has pointed out, some of the unemployed still have jobs and are on temporary layoff. Second, the flow into unemployment of people who have never worked before or who have been out of the labor force seems to be important. This paper ignores both of these categories because it turns out that standard true joblessness of people who have been separated from their jobs is much the dominant source of unemployment in the modern American economy.

With respect to temporary layoffs, Feldstein's tabulations of the March 1974 unemployment survey show that 18 percent of the unemployed are reported as "on layoff"--5 percent with definite recall within 30 days and 13 percent with recall after 30 days or no definite recall. A reasonable guess, supported by Feldstein's other evidence, is that 70 percent of those on layoff are likely to be recalled. Thus 13 percent of all of the unemployed are on temporary layoff, a form of unemployment for which a rather different analysis is appropriate.

In the same tabulation, 25 percent of the unemployed are classified as re-entrants to the labor force. Though there is no direct evidence on the point, it appears that a substantial fraction of the re-entrants have been in the labor force in the very recent past. There is a great deal of difficulty in separating non-workers in the survey into those

who are unemployed and those who are out of the labor force. The survey uses a fairly stringent definition of unemployment which may easily re-classify an unemployed person from the previous month as out of the labor force this month. If such a person is then classified as unemployed in the succeeding month, he will be considered a re-entrant, not a job-loser or leaver. Marston (5) presents evidence on the flows from month to month in the survey that point rather strongly in the direction of very brief spells out of the labor force. According to his data, the average duration of a spell out of the labor force is 8.5 months for adult men and 18 months for adult women, and much lower for teenagers. These averages include all individuals who are permanently out of the labor force because of disability, retirement, or household responsibilities. Those who emerge from the group and become unemployed are very likely to be those who just entered the group. On this basis it appears that a reasonable rough estimate of the fraction of unemployed re-entrants who have been in the labor force within the past six months is 80 percent. Then 20 percent of the re-entrants and all of the new entrants are people whose unemployment cannot be attributed to a recent job separation, a total of 16 percent of the unemployed in 1974.

The three-way breakdown of the unemployed that emerges from this examination of the data for 1974 is:

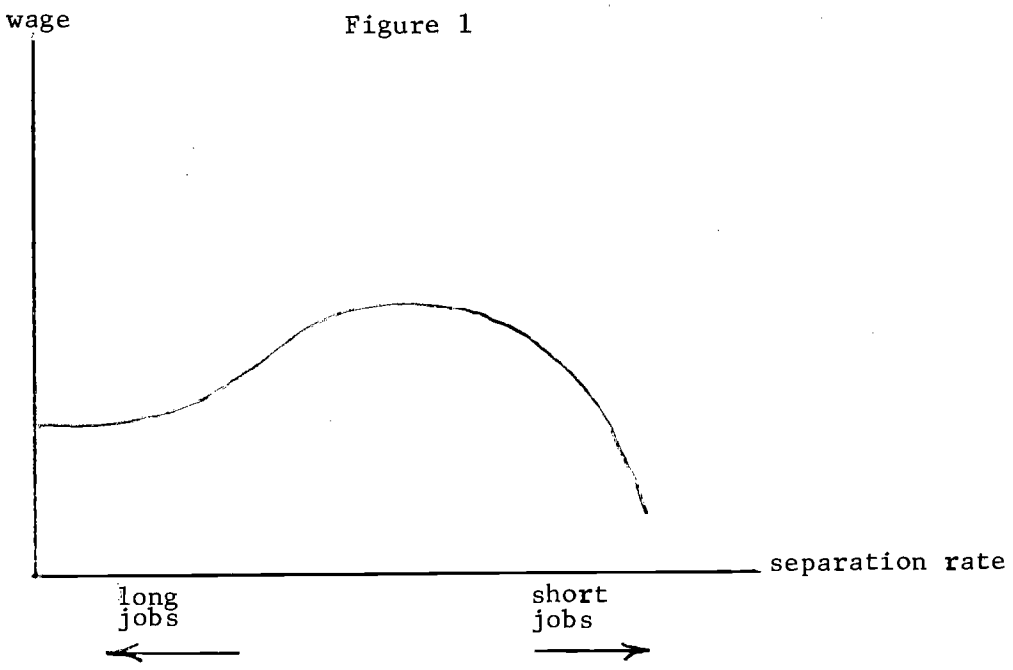
Unemployed but still holding a job	13 percent
Separated from a job recently	71 percent
Previously out of the labor force for at least 6 months	16 percent

Thus the standard view of unemployment applies to nearly three quarters of all of the unemployed. As Feldstein points out, the cyclical fluctuations of temporary layoffs are sharper than the fluctuations of total unemployment, so it is important to study this kind of unemployment to understand disequilibrium. But to understand equilibrium unemployment, the starting point is to study the reasons for job separations. Most of the unemployed got that way by losing or leaving jobs. What is needed to create a theory of equilibrium unemployment is a theory of the impermanence of jobs.

A theory of the duration of employment

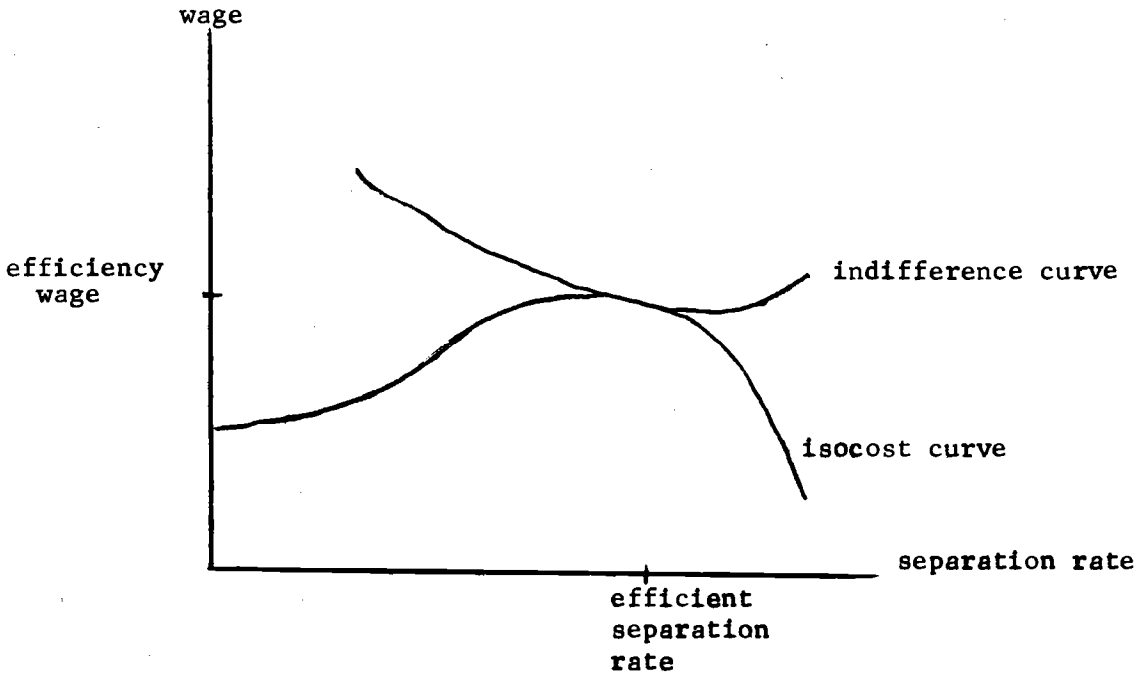
The duration of a job is a matter of concern to both parties to the employment contract. Employers are generally reluctant to agree to very short jobs because of recruiting and training costs. They also find very long jobs costly to offer, because of the implied reduction in the flexibility of their total level of employment. Employers facing product demands that drift over time face very high costs to long-term employment commitments, since there is a substantial probability that the efficient level of employment will fall at some time in the future. In the extreme, small firms facing a probability of bankruptcy simply cannot offer very long or permanent jobs--it is beyond their power to promise not to fail. Employers' views about the duration of employment can be summarized in an isocost curve that permits a higher wage at intermediate durations as against either shorter duration, where turnover costs reduce efficiency, or longer duration, where inflexibility of employment is costly. For reasons that will be made apparent shortly, it is most convenient to plot the isocost curve against the reciprocal of duration, the separation rate:

Figure 1



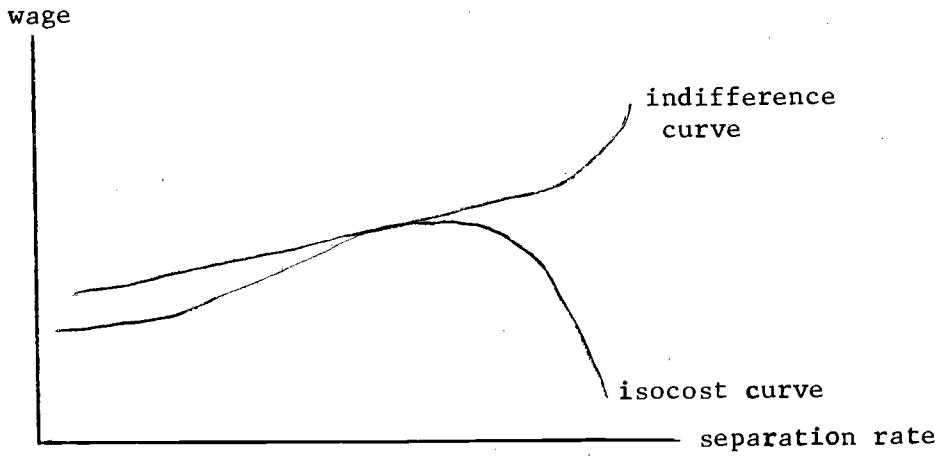
Workers are also concerned about duration. Holding a sequence of brief jobs may be costly because a new job has to be found at the conclusion of each job, and finding work takes time that is uncompensated or compensated at a rate below the wage. For some workers, all job changes are undesirable and, for the same wage, they would always prefer permanent work. Other workers, especially the young, may be willing to buy added flexibility in their lives by choosing briefer jobs. Both considerations can be embodied in an indifference curve, though the slope of the curve is ambiguous. The efficient labor contract between employer and worker specifies a wage and duration (or separation rate) that minimizes cost on the indifference curve, or, equivalently, maximizes workers' satisfaction along an isocost curve. Three cases can be distinguished: First, the efficient point may occur where the marginal rate of substitution between cash income and the separation rate is positive--workers are willing to give up some income to achieve added flexibility:

Figure 2



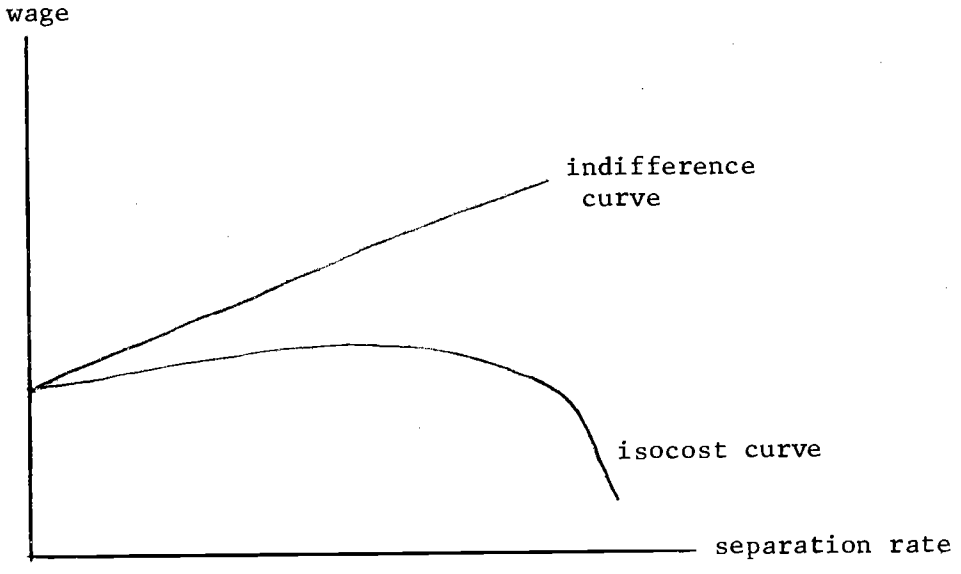
Second, the efficient point may occur where the marginal rate of substitution is negative and workers require higher pay to compensate for shorter jobs and more time spent unemployed:

Figure 3



Finally, the efficient combination may be a corner solution where jobs are permanent:

Figure 4



This corner solution plays an important role in the succeeding discussion. Note that it is impossible to portray it if the duration rather that its reciprocal is on the horizontal axis.

This line of analysis does not actually determine an efficient separation rate and wage, but only an expansion path of alternative efficient combinations. For the time being, however, it will be assumed that there is a perfectly elastic supply of workers to this market provided jobs give a level of satisfaction equal to that available in other labor markets. This restricts the equilibrium to lie along a single indifference curve and so determines both duration and the wage.

How do the parties to the employment contract enforce an agreement about duration? Legal sanctions against quitting a job are weak. Agreements against layoffs are legally enforceable but are not widespread. Designers of employment agreements need to provide the flexibility so that a separation will occur when it is mutually advantageous to both parties (this is the efficiency condition expressed by the tangency in the diagrams) but not permit one party to take advantage of the other. This problem has been discussed extensively in the rather different context of employment contracts where employers insure workers against fluctuations in demand. The present discussion will not attempt much of an answer to the problem, but rather will pursue the implications of contracting over duration in cases where both parties follow the rules after the contract is made. The simplest rule is just to specify the duration of a job as a fixed number of months or years, in which case

the only problem is to enforce a prohibition against quitting, without much assistance from the law. A more efficient procedure is to agree on an expected duration and permit quits or layoffs provided they adhere to the agreed-upon separation probability. Of course, an agreement of this kind is virtually unenforceable, since it will never be clear that any given quit or layoff is a violation. However, a history of layoff rates greater than the promised rate will injure the reputation of the employer, and the same is true for a history of excessive quits on the part of a worker. Thus the inability to enforce the agreement in any one instance does not make it meaningless to agree on expected duration. It seems worth pursuing a theory of efficient duration even though the resulting agreements are not individually enforceable.

The efficient duration of employment depends on the cost of recruiting to the employer and on the cost of finding new jobs to the worker. Tight markets where jobs are easy to find make workers more receptive to shorter jobs and higher separation rates, but impose higher recruiting costs on employers, so employers favor longer jobs. Though the analysis of the efficient duration of employment applies for almost any specification of the operation of the labor market, it seems useful to carry on the discussion within a particular model where it is possible to be completely clear about the mechanics of unemployment and its role in the economy.

In the model to be considered here, there is no private or social value of unemployment in the sense of searching for the best match of worker and job. Jobs and workers are assumed perfectly homogenous. The

unemployed simply form an inventory of workers available for employment. The model also recognizes the pervasive asymmetry of the job-filling process--jobs are filled as soon as they become open, but the unemployed must wait until a job appears for them, which is a stochastic event that may take several periods to occur. The model also assumes that the unemployed accept the first job offered, and that if they receive several offers, they accept one chosen at random.

Suppose that J job offers are made by employers to the U unemployed workers each period. The probability that a particular worker will receive a particular offer is 1/U. The probability that an unemployed worker will receive no offer at all from among the J is

$$1 - f = (1 - \frac{1}{U})^J$$

$$= [(1 - \frac{1}{U})^{-U}]^{-J/U}$$

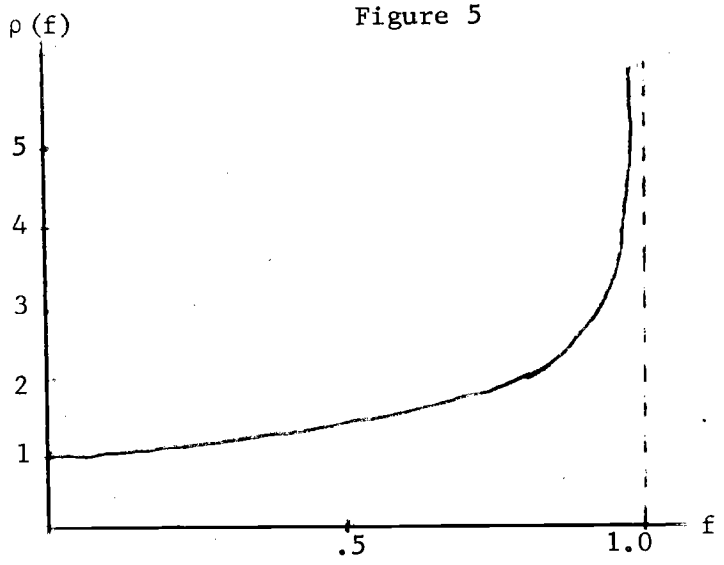
Here f is the rate of job-finding--the probability each period that an unemployed worker will find work. If U is large, the term in square brackets is very close to e, and the job finding rate, f, is

$$f = 1 - e^{-J/U}$$

Since job offers are made at random to the unemployed, some of them may receive more than one offer in one period, and employers must generally make more offers than the number of jobs they hope to fill. Of the J offers made, Uf are accepted. The number of offers needed to yield an expectation of one acceptance is $\rho = J/Uf$. But J/U is functionally related to f: $J/U = -\log(1-f)$, so ρ is just a function of f:

$$\rho(f) = -\frac{\log(1-f)}{f}$$

Figure 5



Recruiting expenses will be assumed proportional to $\rho(f)$--tight markets with f close to one become increasingly costly to employers because many offers must be made to hire a single worker. Note that the benefits of slack markets are almost all available at $f \doteq 0.5$, where $\rho = 1.39$, as against its theoretical lower limit of 1.00. Unless offers are extremely expensive, the socially optimal job-finding rate f will be well above 0.5.

The separation rate s and the job-finding rate f together imply a value for the unemployment rate,

$$u = \frac{s}{s+f}$$

The equations describing the model can now be brought together:

Efficient separation rate:

$$s = \phi(f, \rho(f))$$

Recruiting cost:

$$\rho(f) = -\frac{1}{f}\log(1-f)$$

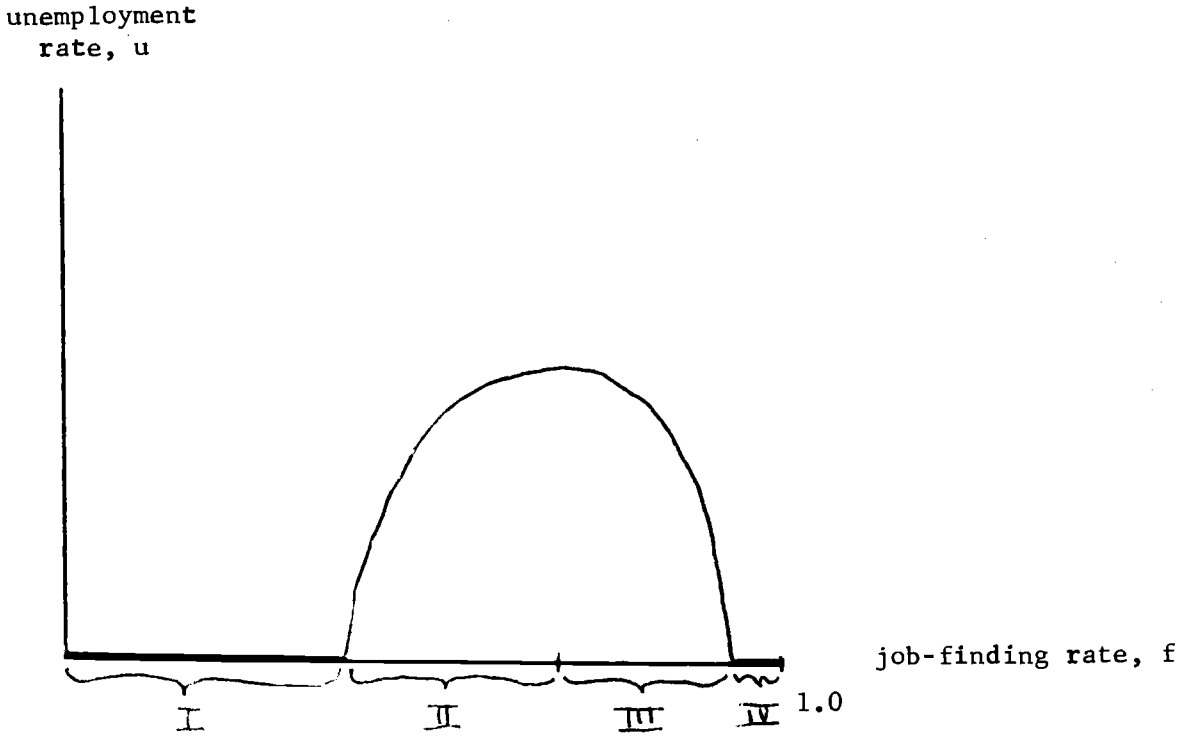
Unemployment rate:

$$u = \frac{s}{s+f}$$

The model has four variables--s, ρ, u, and f, but only three equations. There is a one-dimensional indeterminacy, which is most conveniently indexed by the job-finding rate, f. For any value of f, the model describes a labor-market equilibrium where all the flows into and out of employment and unemployment balance, where no employer can attract workers and also operate at a lower cost than is implied by the equilibrium wage-duration contract, and no worker can find employment at terms different from the equilibrium contract and still achieve a higher level of satisfaction. Yet the equilibrium satisfying all of these conditions is indeterminate.

The implications of the indeterminacy can be seen in the following relationship between the job-finding rate f and the unemployment rate, u:

Figure 6



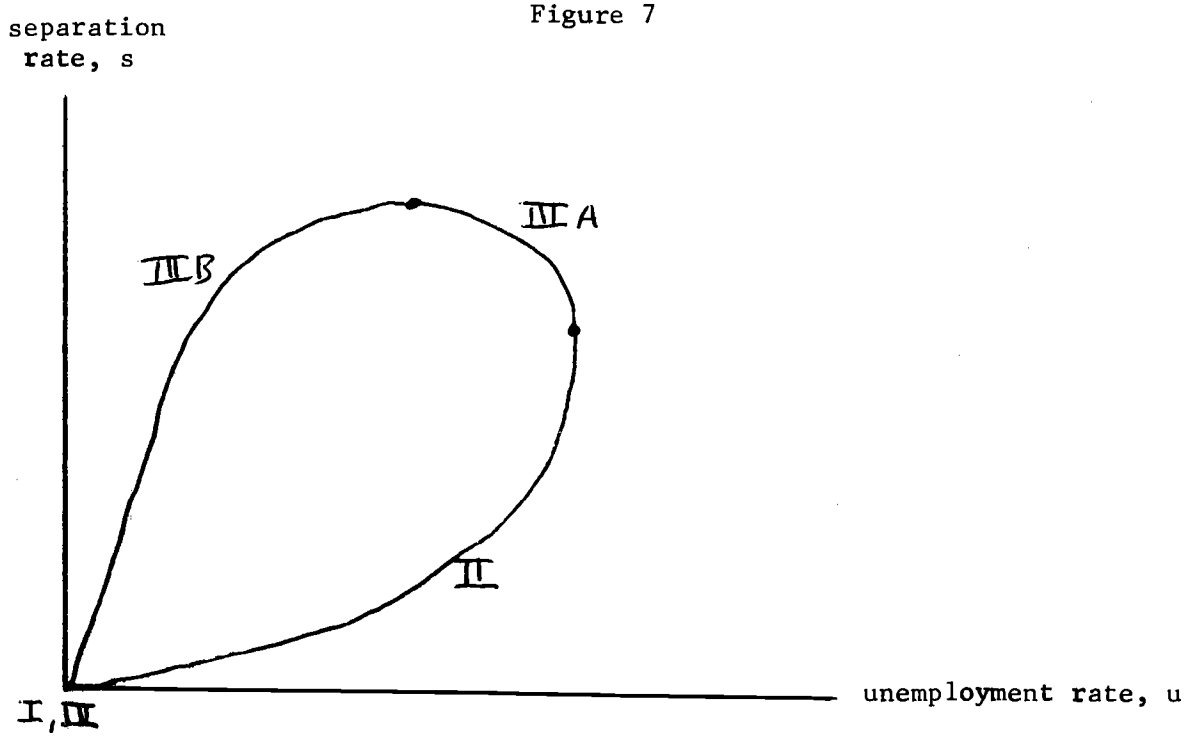
For low values of f, in region I, the cost of the unemployment associated with temporary work is high and there is little compensating benefit to employers in the form of lower recruiting costs, so the efficient duration is infinite--employers offer permanent jobs. The resulting unemployment rate is zero. Participants forego the opportunity to use the labor market to allocate labor dynamically. Of course, the job-finding rate f loses its meaning if there is no unemployment, so the main point of the part of the curve in region I is just that no equilibrium is possible with low job-finding rates and positive unemployment. Depression conditions of extreme unemployment and the near impossibility of finding work cannot be explained by this model. Employers are free to take advantage of such conditions by offering permanent jobs and inducing the unemployed to work for below the prevailing wage in exchange for job security.

For somewhat higher job-finding rates, in the region labeled II, the labor market begins to function as an exchange, so unemployment is positive. Recruiting costs are still unimportant but the efficient duration of employment is high because unemployment is costly to workers. The curve slopes upward in region II because the positive effect of the rising separation rate, s, more than offsets the negative effect of the rising job-finding rate, in the formula for the unemployment rate, $u = s/(s + f)$. A cross section of cities whose labor markets were all in region II would show higher turnover rates, s, in cities with tighter labor markets, as measured by the job-finding rate, f. Further, somewhat paradoxically, cities with tighter markets would have <u>higher</u> unemployment rates.

In region III, unemployment has the more familiar negative relation to the job-finding rate -- tighter markets have lower unemployment. At the boundary between regions II and III where unemployment reaches its highest possible equilibrium value, the separation rate s is still increasing with f. In the left side of region III, the costs of unemployment still dominate so employers respond to a tightening of the labor market by offering shorter jobs. Duration reaches its minimum in the middle of region III. As the market tightens further, the disincentive of recruiting costs becomes important and employers begin to offer longer jobs to limit those costs. Eventually, there is a critical point where the market is so tight that employers find it advantageous to offer permanent jobs even though workers would be perfectly happy to have brief jobs.

The relationship between unemployment and the separation rate implied by the model is the following:

Figure 7



Region I and IV correspond to the same point in the diagram. At this point the labor market does not function to reallocate labor over time, either because workers think that the cost of finding work in the open market is prohibitive or because employers think that the cost of finding workers there is prohibitive. In region II, unemployment and the separation rate are positively related, but both have an unexpected relation to market tightness as measured by the job-finding rate. In region IIIA, unemployment falls as the market tightens, but the separation rate continues to rise. Here it is important to remember that separations include quits as well as layoffs--generally the two have an inverse relation. Finally, in region IIIB unemployment and the separation rate move together and both fall as the job-finding rate rises.

Equilibrium in the long run

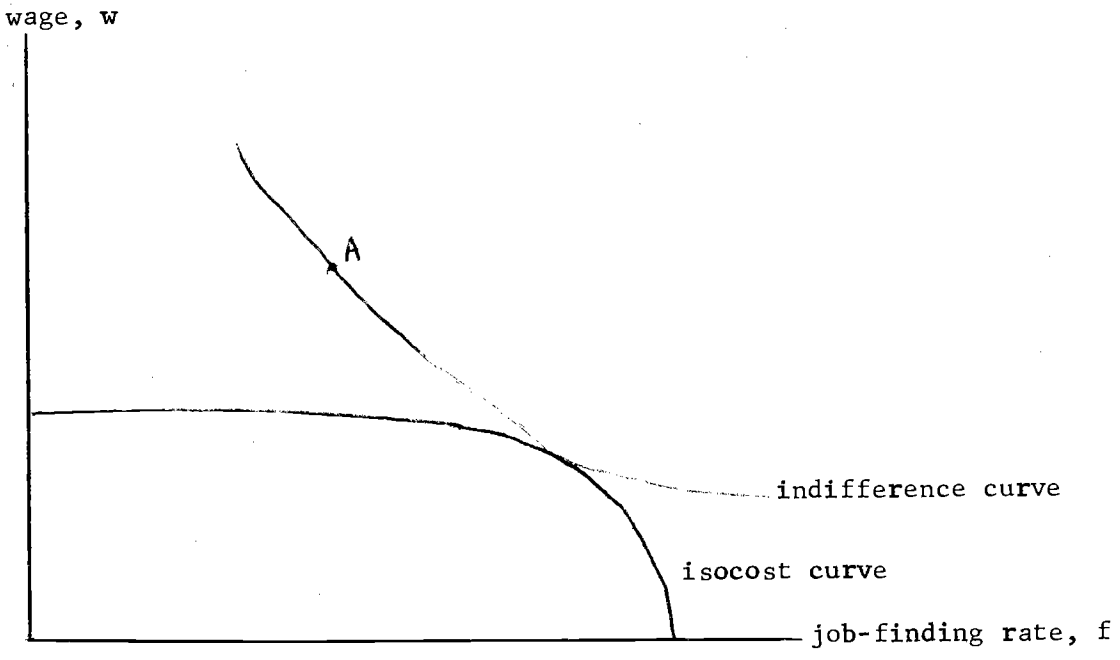
The concept of equilibrium used in the previous section is the conventional requirement that no single participant in the market be able to improve his own situation by changing his own behavior. For example, high unemployment rates cannot be equilibria in the model because individual employers could attract workers at lower wages by offering permanent work. But any combination of job-finding rates and unemployment lying on the curve of Figure 6 offers no opportunity for arbitrage in this sense--the best a single employer can do is to imitate the other employers and sustain the equilibrium.

An equilibrium of this kind may offer unexploited opportunities for profit, however. There is a socially optimal rate of unemployment, and whenever the equilibrium unemployment rate departs from the optimal rate, feasible activities are available that offer pure profit. If these activities are carried out, the optimum becomes the only possible equilibrium, the indeterminacy vanishes, and unemployment need not be a special concern. The activities involve a global intervention in the labor market, though, so the speed and forcefulness of movements toward the optimum may be disappointing. Essentially what is required if the market is too slack, for example, is the creation of a submarket to which workers can be attracted by higher job-finding rates. Employers participating in the sub-market could effectively add the lower unemployment rate to the list of inducements they offer to prospective workers and obtain labor at a lower total cost. To put it another way, the indeterminacy of the equilibrium described earlier

is an implication of the assumption that employers and workers are unemployment-rate-takers. If employers have an instrument for promising workers a more favorable job-hunting experience after the current job ends, they will collectively force the unemployment rate to its optimal level.

Determination of the social optimum involves a balancing of the favorable effects of tight labor markets on the well-being of workers against the costs of congestion in recruiting that tight markets impose on employers. Suppose that the duration of employment is fixed at its socially optimal level. Then the optimal combination of wage payments and job-finding rate occurs at the tangency between the producers' isocost schedule and the workers' indifference curve:

Figure 8



With the duration of employment held fixed, the isocost curve simply reflects the shape of the function $\rho(f) = -\frac{\log(1-f)}{f}$ that determines the number of offers necessary to yield one new hire. The curve is fairly flat until f approaches one, where it begins to fall off rapidly and hits the horizontal axis at a point where all of the cost is consumed by recruiting and none is available for wages. The indifference curve must slope downward, since improved job-finding can only make workers better off when the duration of employment is fixed. In addition to the point of tangency, which is one equilibrium, there is one other equilibrium at a lower value of f and the same value of s, shown as point A. Recall that the separation rate first rises and then falls as a function of f, so for every value of f, there are two possible values of s that are equilibria. Point A is clearly inferior to the optimum, as it is on the same indifference curve but on a higher isocost curve.

Because the only social value of slack markets considered in the model is the increased convenience of recruiting, the social optimum generally occurs at levels of job-finding close to one. For example, suppose time is measured in months, the typical job lasts 25 months, and the cost of making one job offer is 0.1 months of pay. Then the socially optimal rate of job-finding is 92 percent per month and the unemployment rate is close to its technological minimum of 4 percent (all separations generate at least one month of unemployment and the separation rate is 4 percent per month). The model creates the impression that excessively slack markets are more likely than excessively tight ones.

Factors causing movement to the optimum

The model of this paper involves no externalities. An employer who induces a worker to enter a market with unemployment pays the full social cost of the unemployment in the form of a wage premium to the worker. If employers could manipulate the unemployment rate as well as the wage and duration of employment, the social optimum would be the only possible equilibrium. The equilibrium at non-optimal job-finding rates arises because atomisitic employers cannot guarantee their workers a favorable experience in the job market after the present job ends. With respect to a single job and worker, the only instrument available for limiting exposure to adverse conditions in the market is increased duration of employment, and in equilibrium duration has already been set at its best value.

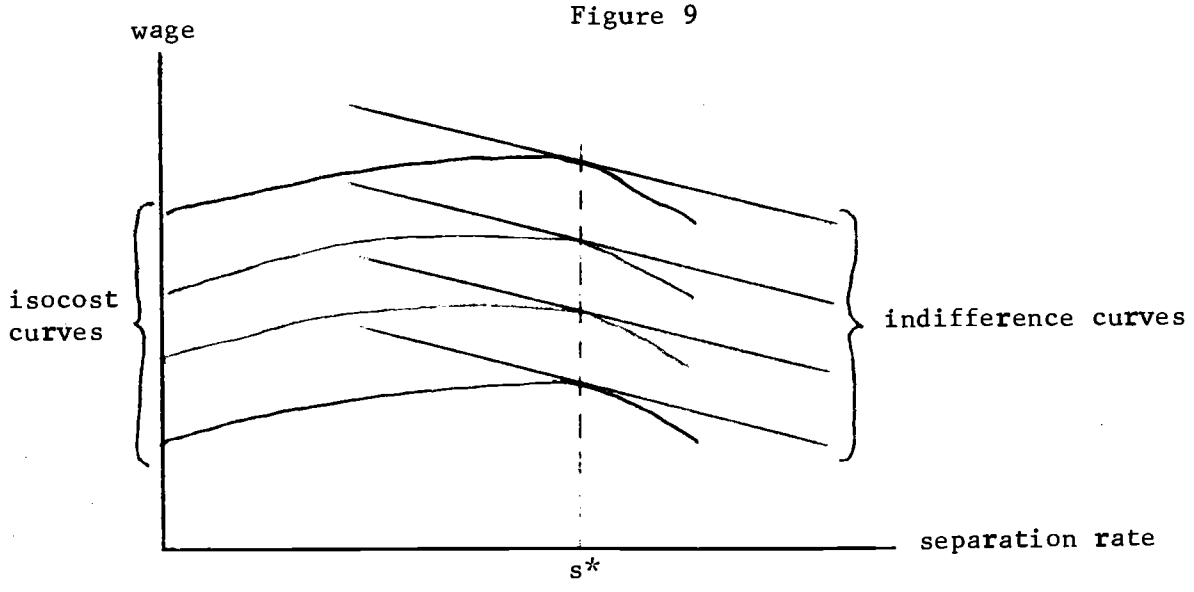
In order to take advantage of the lower cost of labor available at the optimal job-finding rate, an employer or group of employers must create an institution that will replace the existing labor market as a labor exchange. The new market or institution must be large enough to provide the benefits of the continuous a reallocation of labor that make limited duration of employment desirable in the first place, yet it must be controllable so that it does not develop non-optimal job-finding rates itself. Perhaps the simplest such institution is the large diversified firm. Another is the manpower firm that acts as an intermediary in the labor market, offering lower unemployment to its employees and short duration to its customers.

Macroeconomic equilibrium

In the model considered so far, the labor market is embedded in an economy that supplies participants to the market perfectly elastically at a fixed level of utility--all possible equilibria are points on the same indifference surface over wages, duration of work, and job-finding rates. For the labor market as a whole, labor supply is far from perfectly elastic. In fact, a better assumption is that the number of workers available in the aggregate labor market is unaffected by the terms of employment in the market. Still, the previous analysis of efficient employment contracts continues to apply. At whatever level of utility is achieved by workers, the combination of wages and duration should be efficient. Further, there is a socially optimal efficient combination of wages, duration, and job-finding rates, but some doubt about how successfully the market will achieve the optimum.

For the discussion of the aggregate labor market, it will help to make the simplifying assumption that the efficient duration of employment is the same for all levels of worker utility, with the job-finding rate held constant. In other words, the expansion path traced by the tangencies of the indifference curves and isocost curves is a vertical line:

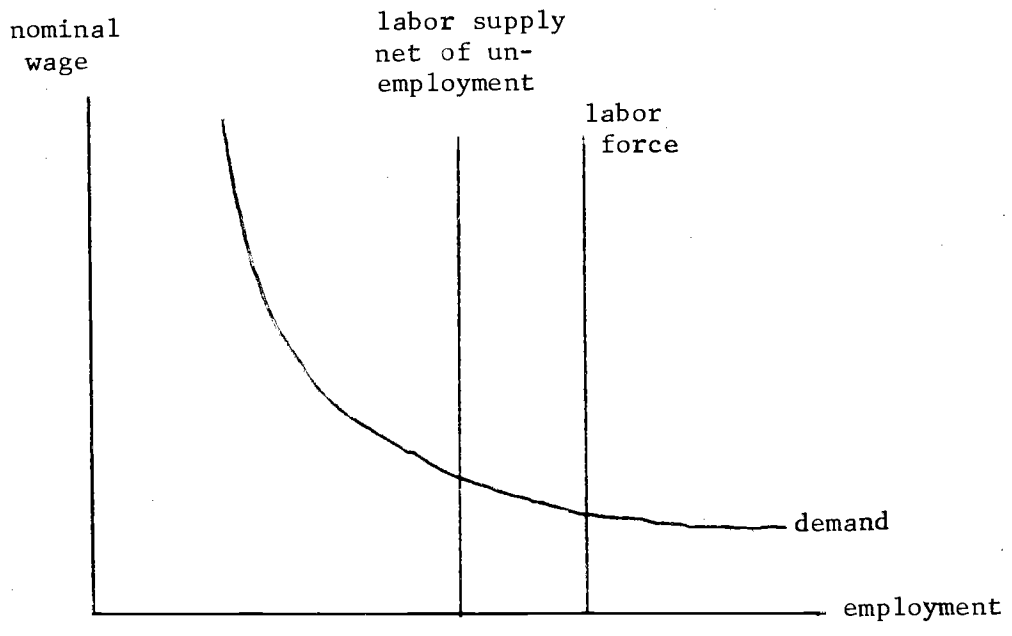
Figure 9



This would be true, for example, if cost is proportional to the wage and if utility is proportional to income. The assumption of a vertical expansion path imposes no important qualifications on the preceding analysis. The efficient separation rate, s*, depends on the job-finding rate, f, just as discussed earlier.

Under this convenient assumption, for a given value of the job-finding rate, the efficient separation rate is independent of the macroeconomic equilibrium, and so is the unemployment rate (recall that $u = s/(s + f)$), Macroeconomic equilibrium is just the classical model adjusted in a minor way to accommodate a predetermined unemployment rate. If the money supply is exogenous, the whole macro-model can be boiled down to a single diagram describing the labor market. Aggregate demand for output translates into a demand function for labor that depends on the nominal wage. The supply of labor, both in the sense of the labor force and the labor force net of the predetermined unemployment, is inelastic with respect to the nominal wage:

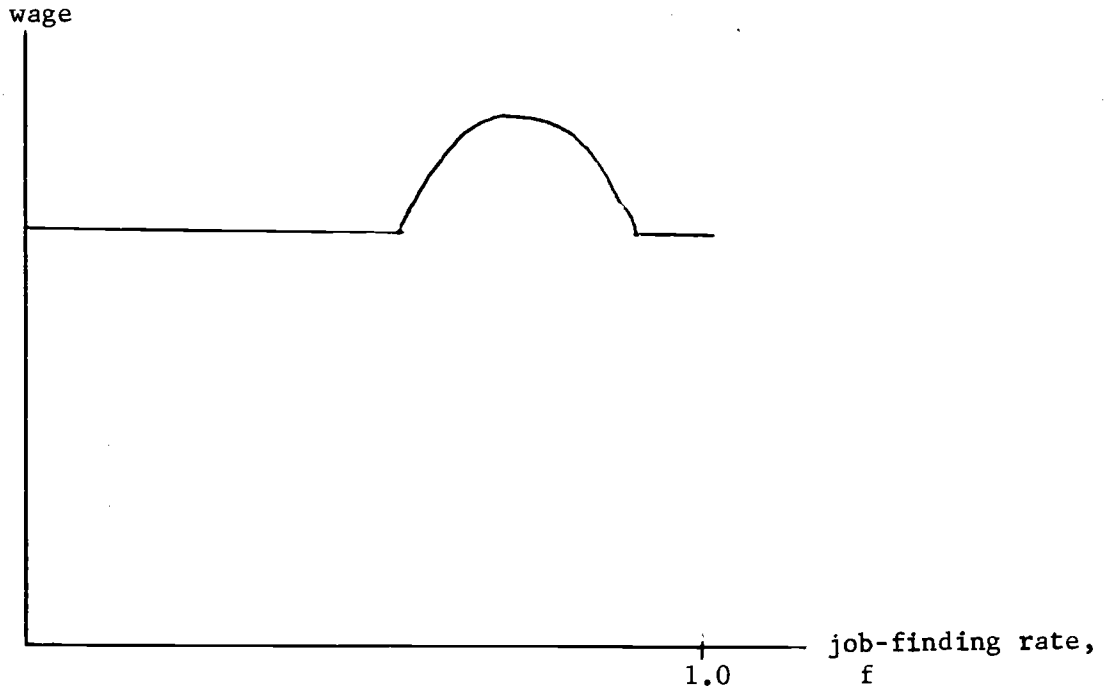
Figure 10



The demand function, whose position is determined by the nominal money supply, among other things, serves to determine the nominal wage, given the predetermined level of employment. Again, this is a classical macroeconomic model. Most economists would agree that to be at all realistic, a rather different short-run dynamic model would have to be added to this to say anything about disequilibrium fluctuations, but this paper is concerned only with equilibrium.

What makes the macroeconomic implications of the present analysis distinctly non-classical is that the job-finding rate is not actually predetermined and the macroeconomic equilibrium is indeterminate in the same sense as in the case of a single labor market. Different job-finding rates correspond to different unemployment rates and thus to different levels of employment and different nominal wages. Recall that the unemployment rate first rises and then falls as a function of the job-finding rate. High unemployment corresponds to low net supply of labor and a higher wage:

Figure 11



the downward-sloping part of this curve seems the most likely and interesting part. Along it, the alternative equilibria are those with higher wages, slacker markets, and higher unemployment as against lower wages, tighter markets, and lower unemployment. As long as f is below its optimal value, real incomes are higher in tight markets even though nominal wages are lower, for two reasons: First, wages are received for a larger fraction of the year when the labor market is tight and unemployment is low. Second, workers are more productive when markets are tight because the more fluid labor market lets employers make better use of their workers by making job separations less costly.

In the aggregate labor market, as in individual markets, there are longer-term forces moving toward the optimal unemployment rate, but these forces are not the result of the self-interest of small participants in the market. If the economy finds itself in the stagnant condition described by the slack-market equilibrium, it may move out of it only very slowly. Note that the tightening of the market moves against the disequilibrium wage-unemployment relation described by the Phillips curve--lower unemployment means lower nominal wages.

Concluding remarks

A model with indeterminate labor-market conditions is incapable of answering the question of how the economy came to be where it is. Tight or slack markets are a matter of accident in the model. Only a dynamic version of the model could address the issue of what moves the job-finding rate in the short and medium runs. Nothing in the model rules out the possibility that unexpected increases in demand can cause the market to tighten, though it is equally true that the equilibrium job-finding rate might be unaffected by such an increase. Obviously the model does not imply any simple policy frontier of unemployment against inflation to guide the choice of aggregate policy. On the one hand, the considerations of this paper greatly complicate the analysis of aggregate policy-making. On the other hand, the analysis suggests some relief from the terribly pessimistic conclusions of contemporary Phillips curves, which suggest that the restoration of tight labor markets would be accompanied by high and accelerating inflation. A tight-market equilibrium with no worse inflation may be available today, though this paper makes no suggestions about how to get to it.

References

(1) Robert J. Barro, "Unanticipated Money Growth and Unemployment in the United States", American Economic Review, vol. 67, pp. 101-115, March 1977.

(2) Martin Feldstein, "The Importance of Temporary Layoffs: An Empirical Analysis", Brookings Papers on Economic Activity,1975:3, pp. 725-744.

(3) Robert E. Hall, "Turnover in the Labor Force", Brookings Papers on Economic Activity, 1972:3, pp. 709-756.

(4) _____, "Expectation Errors, Unemployment, and Wage Inflation", Center for Advanced Study in the Behavioral Sciences, September 1977, processed.

(5) Stephen T. Marston, "Employment Instability and High Unemployment Rates", Brookings Papers on Economic Activity, 1976:1, pp. 169-203.

(6) Thomas J. Sargent, "Rational Expectations, the Real Rate of Interest, and the Natural Rate of Unemployment", Brookings Papers on Economic Papers on Economic Activity, 1973:2, pp. 429-472.

(7) _____, "A Classical Macroeconometric Model for the United States", Journal of Political Economy, vol. 84, pp. 207-238, April 1976.

(8) _____, "The Persistence of Aggregate Employment and the Neutrality of Money", University of Minnesota, January 1977, processed.