A THEORY OF THE NATURAL UNEMPLOYMENT RATE
AND THE DURATION OF EMPLOYMENT*

Robert E. HALL
National Bureau of Economic Research, Stanford, CA 94305, USA

Received November 1978

In this paper, a theory of the natural or equilibrium rate of unemployment is built around a theory of the duration of employment. Evidence is presented that most unemployed workers became unemployed because their previous jobs came to an end; only a minority are on temporary layoff or have just entered the labor force. Thus, high-unemployment labor markets are generally ones where jobs are brief and there is a large flow of newly jobless workers. The model of the duration of employment posits that employment arrangements are the efficient outcome of the balancing of workers' and employers' interests about the length of jobs. Full equilibrium in the labor market also requires that the rate at which unemployed workers find new jobs be efficient. The factors influencing the resulting natural unemployment rate are discussed. Under plausible assumptions, the natural rate is independent of the supply or demand for labor. Only the costs of recruiting, the costs of turnover to employers, the efficiency of matching jobs and workers, and the cost of unemployment to workers are likely to influence the natural rate of unemployment strongly. Since these are probably stable over time, the paper concludes that fluctuations in the natural unemployment rate are unlikely to contribute much to fluctuations in the observed unemployment rate.

1. Introduction

Milton Friedman's (1968) famous definition of the natural unemployment rate runs as follows:

The 'natural rate of unemployment' ... is the level that would be grounded out by the Walrasian system of general equilibrium equations, provided there is imbedded within them the actual structural characteristics of the labor and commodity markets, including market imperfections, stochastic variability in demands and supplies, the cost of gathering information about job vacancies and labor availabilities, the costs of mobility, and so on.

*This research was supported by the National Science Foundation and is part of the NBER's program of research on Economic Fluctuations. The paper has not undergone the review accorded official NBER publications; in particular, it has not been submitted for approval by the Board of Directors. The author is grateful to Robert Barro for helpful comments.
Though this definition is hardly more than a list of things to think about in creating a theory of the natural rate, the basic notion that there is a natural rate has become reasonably well established, even among economists who resisted the idea at first and who now accept the principle but give it another name. Most recent theoretical work on unemployment has focused on the search activities of the unemployed or on the unemployment of workers with permanent employment contracts who are on temporary layoff. But most of the unemployed have truly lost their jobs, and existing theories have little to say about why they became unemployed in the first place. By and large, workers become unemployed because their jobs are not permanent. The shorter the duration of employment, the greater is the flow of workers into the pool of the unemployed.

In this paper, a theory of the natural rate of unemployment is built around a theory of the duration of employment. The paper does not concern itself with disequilibrium or with problems of lags in the adjustment of wages, and so is not aimed at deriving a Phillips curve. On the contrary, it is concerned only with the unemployment that exists in a labor market in equilibrium. Some, but not all, of the considerations in Friedman's list are incorporated in the theory. The spirit of Friedman's definition is preserved – unemployment is treated as a phenomenon that can be understood within a general equilibrium Walrasian model, provided the model is suitably extended.

In creating a theory of unemployment in equilibrium, one must start with more than the proposition that unemployment is the difference between supply and demand in the labor market, though this proposition is always true. The paper begins with a basic, general definition of equilibrium: In equilibrium, no participants in the market have unexploited opportunities to make themselves better off. At the equilibrium unemployment rate, employers cannot obtain labor at lower cost by offering work to the unemployed at below the market wage. 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. The problem considered here is the next step beyond the one treated in the recent literature on fixed-price equilibria. A fixed-price equilibrium is not an equilibrium at all in the sense the term is used here – opportunities for self-improvement abound in a fixed-price equilibrium, as long as the possibility of agents offering to transact at prices different from the fixed prices is recognized.

Before the development of the model of the duration of employment, it is useful to mention some evidence on the quantitative importance of job separations as a source of unemployment. There are two other principal sources of unemployment apart from separations: temporary layoffs and entrance to the labor force. In spite of the attention that temporary layoffs have received recently as a result of the work of Martin Feldstein (1975), they are not a major source of unemployment in the contemporary American economy. In 1974, only 18 percent of all unemployed workers were reported as 'on layoff', and only about 70 percent of them, or 13 percent of total unemployment, could reasonably expect to be recalled to their old jobs. Further, it appears that only 16 percent of the unemployed have not worked in the six months before becoming unemployed and so cannot attribute their unemployment to a recent job separation. The remaining 71 percent of the unemployed consist of people who have become unemployed by irrevocable separation from their earlier jobs. 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 is to create a theory of equilibrium unemployment is a theory of the impermanence of jobs.

In the model developed in this paper, 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 tangente 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.

2. A model of unemployment

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 homogeneous. The unemployed simply form an inventory of workers available for employment. The model also recognizes the pervasive asymmetry of the job-filling process — employers fill jobs much more rapidly than workers find jobs. In the model, 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 \( U \) job seekers each period. The probability that a particular worker will receive a particular offer is \( 1/U \). The probability that a job seeker will receive no offer at all from among the \( J \) is

\[
\]

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, some job seekers 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 \( \rho \) is just a function of \( f \) (see fig. 1).

\[
\rho(f) = -\log(1 - f)/f.
\]

Recruiting expenses will be assumed proportional to \( \rho(f) \) — tighter markets with \( f \) approaching zero 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 \( f = 0.5 \), where \( \rho = 1.39 \), as against its theoretical lower limit of 1.00.

Now let \( s \) be the separation rate, that is, the fraction of the employed who become job-seekers each period. Together with the job-finding rate, \( f \), it implies a value for the unemployment rate when the market is in stochastic equilibrium, defined as equal flows into and out of unemployment. Suppose

\[
U = (1 - f)U + sE,
\]

or

\[
U/E = s/f.
\]

Now the unemployment rate, defined as the fraction of the labor force that is unemployed throughout the period, is

\[
u = \frac{(1 - f)U}{E + (1 - f)U} = \frac{s}{s + f/(1 - f)}.
\]

The quantity \( f/(1 - f) \) is the number of jobs found by all job-seekers for each person who remains unemployed throughout the period.

The next section of the paper examines the determinants of the separation
rate, \( s \), and the subsequent section does the same for the job-finding rate, \( f \). Then the two are combined to form a model of the equilibrium or natural unemployment rate.

3. 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.\(^2\) 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.

![Diagram](image)

Fig. 2

\(^2\)The first of these considerations has long been emphasized in the literature on labor turnover; see Pencavel (1972) and the many references he cites.

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 minimum cost along 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.

![Diagram](image)

Fig. 3

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, as shown in fig. 4. Finally, the efficient combination may be a corner solution where jobs are permanent, as in fig. 5.

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 analysis so far has been carried out in terms of tangencies of the iso-cost and indifference curves, so it determines an expansion path of alternative efficient combinations of wages and separation rates. The position of the workers’ indifference curve depends on the cost of finding new jobs, indexed by the job-finding rate, $f$. The shape of the firms’ iso-cost curve depends on recruiting cost, $p(f)$, which in turn depend on $f$. The tangency of the two curves at a given level of worker satisfaction, say $y$, can be described by a function of $f$ and $y$,

$$s = \phi(f, y).$$

There is another function of $f$ that gives the wage at the point of tangency, but it will not be needed in what follows.

4. The efficient job-finding rate

In the previous section, the job-finding rate was held constant at an arbitrary level and the considerations entering the determination of the efficient separation rate were explored. This section does the reverse. The separation rate is held constant and the efficient job-finding rate is derived. Again, efficiency involves a tangency of an iso-cost curve to an indifference curve. Both wages and job-finding are goods to the consumer and bads to the producer, so the two curves have conventional shapes, as shown in fig. 6. The set of tangencies traces out an expansion path of alternative efficient job-finding rates given the separation rate,

$$f = \psi(s, y).$$

Again, the expansion path is indexed by the level of satisfaction, $y$, achieved by workers. Each point on it represents the balancing of the marginal costs to employers of congestion in the labor market associated with higher job-finding rates against the marginal benefits to workers of lower unemployment.
Then the efficient job-finding rate minimizes cost subject to the constraint of achieving a given effective income, $y$,

$$\min \int w\lambda(s)(r\rho(f) + 1),$$

subject to

$$\int \frac{f}{(1-f)s + f} w = y,$$

or

$$\min \int \left(1 + s\frac{1-f}{y}\right)w\lambda(s)(r\rho(f) + 1).$$

But this is equivalent to

$$\min \int \left(1 + s\frac{1-f}{f}\right)(r\rho(f) + 1),$$

which involves no unknown functions. The minimizing $f$ is the value of the function $\psi(s, y)$ defined earlier and does not depend on $y$. Some representative values of $\psi$ are presented in table 1. For a given value of the recruiting cost, $r$, the efficient job-finding rate is almost unchanged across a wide range of separation rates, $s$. Consequently, almost all of the differences in unemployment rates reflect differences in separation rates, not in job-finding rates. To put it another way, unemployment rates differ because the frequencies of unemployment differ, not because the durations differ.

Even without the special separability assumptions that make it possible to compute $\psi(s, y)$ exactly, it seems reasonable to expect that $\psi(s, y)$ will not depend very much on either of its arguments. The efficient job-finding rate is

\[ y = (1-u)w = \frac{f}{(1-f)s + f} w. \]

\[ u = \frac{s}{s + f/(1-f)}. \]

---

*The function $\lambda(s)$ is the outcome of the optimal staffing policy given the separation rate. It will depend on the stochastic properties of demand and other considerations not treated explicitly here.*
determined primarily by the balance of search costs against recruiting costs. Both are roughly proportional to wages, so the balance is not sensitive to the level of satisfaction achieved by workers. Similarly, both the marginal benefit and cost of a high job-finding rate are roughly proportional to the separation rate, so it has little effect on the balance as well.

5. Equilibrium unemployment

Equilibrium in the labor market requires first that the number of people looking for work balance the number of people who want to work, net of those who are in the process of finding work, and second that the terms of employment be efficient. Conditions in the market are measured by the wage rate, \( w \), the separation rate, \( s \), and the job-finding rate, \( f \). The demand for labor is a function of these three variables: \( L^D(w, s, f) \). Similarly, the number of workers attracted to the market is a function of the same variables, \( L(w, s, f) \). However, a fraction, \( u \), of the workers will be looking for work at any one instant, so the appropriate equilibrium condition is

\[
L^D(w, s, f) = (1 - u)L(w, s, f),
\]

or

\[
L^D(w, s, f) = \frac{f}{(1-f)s + f} L(w, s, f).
\]

The level of satisfaction achieved by workers is also a function of the employment terms,

\[ y = g(w, s, f). \]

The analysis presented earlier in the paper characterized the conditions of efficient employment in terms of the two functions,

\[ s = \phi(f, y) \quad \text{and} \quad f = \psi(s, y). \]

Given a wage level, \( w \), the utility function \( g(w, s, f) \) and the two equations can be solved to get the efficient separation rate and job-finding rate as functions of \( w \).

\[ s = s^*(w) \quad \text{and} \quad f = f^*(w). \]

There are good reasons to believe that neither \( s \) nor \( f \) is very sensitive to \( w \) under the assumptions of separability introduced in the previous sections, \( s^* \) and \( f^* \) are in fact just constants, independent of \( w \).

Equilibrium in the market can now be expressed in a single equation,

\[
L^D(w, s^*(w), f^*(w)) = \frac{f^*(w)}{(1 - f^*(w))s^*(w) + f^*(w)} L(w, s^*(w), f^*(w)),
\]

or in a diagram.

Equilibrium occurs at the intersection of the demand schedule \( L^D(w) \) and the net supply of labor, \( (1 - u(w))L(w) \). The level of employment, \( E \), is the horizontal coordinate of this point, and the labor force \( L \) is the point on the labor supply schedule, \( L(w) \), at the equilibrium wage.

Under mild assumptions about the smoothness of the various functions, an equilibrium is guaranteed to exist in the labor market. Beyond this rather general statement, not much can be said about the equilibrium without introducing further assumptions. The separability conditions discussed in the previous section turn out to give some much more definite results. In the first place, when both costs and utility are proportional to the wage, the efficient separation rate and job-finding rate are independent of the level of worker satisfaction. The efficient separation rate, \( \phi(f) \), is obtained from

\[
\min_{s} \left(1 + s \frac{1-f}{f}\right) \lambda(s)(r \phi(f) + 1),
\]
and depends only on \( f \), the occupancy rate, \( \lambda(s) \), and the recruiting rate, \( r \).

The efficient job-finding rate, \( \psi(s) \), is obtained from

\[
\text{Min}_f \left( 1 + s \frac{1-f}{f} \right) (rs\phi(f) + 1),
\]

and depends only on \( s \) and the recruiting cost, \( r \), as discussed earlier. The intersection of the two schedules determines the efficient levels of \( s \) and \( f \), and so determines the efficient unemployment rate, quite independently of the equilibrium levels of employment and the labor force. Under the separability assumptions, the natural unemployment rate is unaffected by shifts in either the labor supply or labor demand functions.

It is possible (but tedious) to demonstrate the following propositions about \( \phi(f) \) and \( \psi(s) \):

1. The efficient separation rate, \( \phi(f) \), is an increasing function of the job-finding rate, \( f \) – jobs are easier if they are easier to find.
2. The efficient separation rate is a decreasing function of the recruiting cost, \( r \).
3. The efficient separation rate increases if the occupancy rate, \( \lambda(s) \), shifts so as to lower the marginal cost of a higher separation rate.
4. The efficient job-finding rate, \( \psi(s) \), is an increasing function of the separation rate – if jobs last longer, they should not be as easy to find.
5. The efficient job-finding rate is a decreasing function of the recruiting cost.
6. The efficient job-finding rate is unaffected by shifts in the occupancy rate, \( \lambda(s) \).

As shown in the previous section, the efficient job-finding rate, \( \psi(s) \), is very insensitive to \( s \). It appears that the efficient separation rate, \( \phi(f) \), is similarly insensitive to \( f \), though this depends on the curvature of the occupancy rate function, \( \lambda(s) \). If so, one schedule is nearly horizontal and the other is nearly vertical, as shown in fig. 8. The efficient combination is at the intersection of the two schedules, \( f^* \) and \( s^* \), but \( f^* \) is much higher by the \( \psi(s) \) and \( s^* \) by \( \phi(f) \). An increase in recruiting costs shifts \( \phi(f) \) to the left and \( \psi(s) \) downward, so both \( f^* \) and \( s^* \) decline. Turnover declines, but the effect on the natural unemployment rate is ambiguous. A change in the occupancy rate that shifts \( \phi(f) \) to the right has no effect on \( \psi(s) \), so almost all the effect is to increase the separation rate. The natural unemployment rate increases.

Evidence from cross sections has tended to show that most differences in unemployment rates are associated with differences in separation rates, not in job-finding rates [Hall (1972)]. In terms of the model presented here, such variations are more likely to reflect differences in the costs of turnover to employers, as measured by \( \lambda(s) \), rather than in recruiting costs, \( r \). Recall that \( r \) is stated as the number of periods of wages, so it automatically makes recruiting costs proportionately higher for workers with higher wages.

6. Concluding remarks

In the model of this paper, the natural unemployment rate is the outcome of efficient employment arrangements. When the labor market is in equilibrium and unemployment is at its natural level, no alternative combination of wages, job duration, and job-finding rates could make workers better off without also increasing costs to employers. If the matching process embodied in the model is the best that can be done to bring workers and employers together, then the natural rate just described is also the social optimum. There are no externalities in the model, and no good case for government intervention in the labor market. It appears that the socially optimal, natural unemployment rate is quite low. The only influence limiting the tightness of the labor market is the cost of congestion in recruiting that tight markets impose on employers. This is important only as the job-finding rate approaches one.

How strong are the economic forces that push the labor market toward its equilibrium? A full answer requires a theory of the market in disequilibrium, which exceeds the ambitions of this paper. A few comments do seem
indicated by the equilibrium analysis, however. Of the three components of the employment bargain, two—the wage and the separation rate—are under the direct control of the firm. It is hard to see how a disequilibrium could persist where firms had an unexploited opportunity to reduce costs by changing these two terms. But this does not imply a speedy movement to full equilibrium. Unemployment could be far above the natural rate, and yet the tradeoff between wages and the duration of employment could be efficient. This will occur if the job-finding rate is below its efficient level. Unemployment will be high, but the wage-duration terms must be more favorable than their equilibrium values in order to attract workers, because they know how hard it is to find work. In order to lower wage costs, firms must offer prospective workers better chances of finding jobs than prevail in the market at large. If employers are large and can communicate the existence of job openings effectively to job-seekers, then they may be able to attract workers at the fully efficient terms. Smaller firms who recruit more passively cannot do this by themselves, but some kind of intermediary could establish a private labor market where conditions could be regulated to maintain efficiency. Relative to a slack general market, wages would be lower, and so would labor costs, but workers’ incomes would be as high as in the general market because work would be easier to find. Probably the easiest way would be to prescribe one or both of the terms of employment (wage and duration). This is not altogether different from the activities of firms offering temporary clerical help in the United States today.

Simple unilateral activities of individual firms or workers do not move the labor market quickly to the point of efficient employment terms and to the natural unemployment rate. The types of activity that achieve the efficient job-finding rate take time and resources, so there is every reason to believe that a market could spend long episodes with unemployment above or below the natural rate.

The analysis of this paper seems to point in the direction of a natural unemployment rate that is largely unaffected by general economic forces. Under plausible assumptions about costs and preferences, the natural rate in a market is independent of supply and demand or the wage level in the market. It depends only on certain relative costs that seem likely to remain stable over time. Fluctuations in the natural rate itself seem unlikely to be an important part of cyclical fluctuations in unemployment. Rather, the analysis suggests that fluctuations in labor market conditions of a sort that are difficult for individual firms to offset are the most promising explanations of the facts.

References


