

ROBERT E. HALL

*Stanford University*

*and National Bureau of Economic Research*

# *Employment Fluctuations and Wage Rigidity*

DURING the past decade, two facts about the U.S. labor market became more apparent than ever before: the large magnitude of fluctuations in employment and the lack of any strong response of wages to these fluctuations. The year 1975 saw the most striking manifestations of these features. Total labor input to the private economy fell by 6 percent from a year earlier (relative to trend growth), while wage inflation continued at close to its rate in the preceding boom. Although macroeconomists have puzzled over these characteristics ever since the discipline came into being, efforts redoubled in the 1970s to provide a solid economic rationale for the insensitivity of wages to current economic conditions and for the conspicuous deviations of employment from the smooth trend predicted by simple theories of economic growth.

Ten years ago macroeconomists were satisfied with a simple idea that had become virtually the ruling doctrine after Keynes—money wages are predetermined, or at least are quite unresponsive to current economic conditions. Firms set employment unilaterally by hiring up to the point where the marginal revenue product of labor equals the sticky wage. If nominal aggregate demand falls, employment falls. This idea inhabits every textbook in intermediate macroeconomics and underlies much professional analysis. But a defect in this line of thought has been apparent for many years, and has become more of an embarrassment to macro-

The author is grateful to the National Science Foundation for financial assistance. I thank James L. Medoff, Ben S. Bernanke, and members of the Brookings panel for helpful comments, and Ben Craig for assistance.

economics as the field has drawn closer to microeconomics: whenever inadequate demand pushes employment below its market-clearing level, economic inefficiency results. If workers and employers could get together and agree on the level of employment, they would equate the marginal revenue product of labor not with the wage but with the marginal value of workers' time. Employment would not be distorted by a sticky money wage. Demand and supply would have equal roles in employment determination, instead of the predominance of demand as in traditional macro theory.

Serious investigation of the idea that there are better ways for workers and employers to deal with each other as aggregate demand varies has led in a number of directions. In order to understand most of the new ideas, it is important to keep in mind another fact about the U.S. labor market—most workers hold jobs for quite a few years. Employers and workers typically have long-term relations with each other. One of the most significant lines of recent thought pursues the implications of this important fact. Wages are insensitive to current economic conditions because they are effectively installment payments on the employer's obligation to transfer a certain amount of wealth to the worker over the duration of the employment arrangement. A major corollary is the limited allocational role of the wage payment for employment. The rule of the open market—set the value of the marginal product of labor equal to the current wage—no longer has meaning when the current wage is a more or less arbitrary payment on a long-term obligation. Instead, the more fundamental principle of equating the marginal revenue product to the marginal value of labor's time should govern. This basic condition of economic efficiency is the starting point for recent thought on employment fluctuations within long-term employment arrangements.

In this paper much of the discussion is devoted to the issue of employment efficiency. It is one thing to argue that employment arrangements at the level of the individual firm result in an efficient flow of labor services from one worker to that firm, and quite another to argue that the total flow of labor services from all workers to the aggregate economy is efficient. What I call the micro efficiency condition requires that the employment level equates the marginal product of labor with the marginal value of time; it seems to explain a lot about the institutional arrangements for employment determination. The macro efficiency condition is much more

ambitious, requiring that every worker be in a job that makes the best use of the worker's time. Macro efficiency is refuted every few years by recessions, when total labor input falls and numerous workers spend weeks or months out of work. Macro efficiency is promised by the theory of competitive markets, but free economies periodically fall a little short of the theoretical ideal. Much remains to be understood about the failure of macro efficiency in an economy in which individual agents try hard to achieve micro efficiency in their own employment arrangements.

The relative importance of on-the-job adjustments in labor input compared to adjustments from movements among jobs is revealed in the data on output and employment. The paths of these variables have been anything but smooth, especially during the past decade. Fluctuations in output have been larger proportionally than fluctuations in the total volume of work, measured as employee hours. Within the theory of long-term employment arrangements, this reflects the operation of an implicit or explicit agreement that employees work harder when there is more work to do. Similarly, there are important cyclical fluctuations in annual hours of work and, again, these are interpreted as the working of the micro efficiency principle. On the other hand, data on wage movements show relatively little variation over time even though important variations in the demand for labor seem to have taken place. The theory of long-term employment arrangements points to the installment-payment nature of wages to explain their unresponsiveness.

Although the micro efficiency principle appears to be helpful in understanding some of the cyclical movements of employment, the macro efficiency hypothesis is an untenable generalization for the U.S. economy. In the aggregate, efficiency requires the equality of the marginal rate of substitution and the marginal rate of transformation between goods and working. What is lacking is a convincing explanation for sharp cyclical contractions in output that is consistent with this efficiency condition. Neither the public's preferences about work and consumption nor the productive technology shift suddenly, as far as can be determined. Economists may acknowledge that people work harder when there is more work to do, but the macro efficiency principle does not explain why there is sometimes distinctly less work to do in the whole U.S. economy.

A second and closely related defect of the macro efficiency hypothesis is the apparent vulnerability of aggregate output to purely monetary

shocks. The efficiency conditions that ought to determine employment are exclusively real and should be unaffected by shifts in the money stock. This proposition is unambiguously refuted by data for the United States.

Third, theories based on long-term employment arrangements and the efficient use of workers' time leave much unexplained about the level of and changes in unemployment in the U.S. economy. Although the definition of unemployment used in the United States classifies permanent job-holders on temporary layoff as unemployed, they form only a tiny fraction of total unemployment except in times of rapid contraction of the economy. In normal times, the most unemployment seems to occur among groups who work relatively little; if the level is efficient, it means that their time is better spent at home and in other activities outside the labor market. Even in recessions, most of the increase in unemployment is among workers who have unambiguously lost jobs and are looking for new ones.

### **Cyclical Movements of Output, Employment, and Wages**

Table 1 shows some basic puzzles in the data on the aggregate labor market. Column 1 indicates that total labor input (measured as total hours of work) grew at an average of 1.5 percent a year, but the growth was far from smooth. In both recessions of the decade, labor input fell dramatically, by 5.2 percent below trend in 1970–71 and by 6.6 percent below trend in 1974–75. In the other years, especially 1973, 1977, and 1978, labor input grew much faster than trend. The total amount of work done by a typical member of the working-age population fluctuated greatly. Part of these fluctuations show up as complementary movements of the unemployment rate, as shown in column 2 of the table. Column 3 displays another major puzzle—the almost complete absence of market-clearing movements of wages. Wages continued to rise smoothly during the decade. The two recessions, each having large reductions in labor input and increases in unemployment, were accompanied by only slight moderations in wage inflation.

One of the major points of this paper is the importance of long-term jobs in understanding the puzzles of the behavior of employment and wages. Table 2 elaborates on the labor input part of table 1 by showing the movements of its various components. Both columns 2 and 3 reflect the extra work done by employees when there is more work to do. The well-known cyclical behavior of productivity, shown in column 2, is an

**Table 1. Labor Input, Unemployment, and Wage Inflation, 1970–79<sup>a</sup>**  
Percent

<i>Year</i>	<i>Rate of change of total hours of work (1)</i>	<i>Unemployment rate (2)</i>	<i>Rate of change of wages (3)</i>
1970	−1.6	4.9	7.1
1971	−0.6	5.9	6.7
1972	3.0	5.6	6.3
1973	3.9	4.9	8.2
1974	0.7	5.6	9.1
1975	−4.3	8.5	9.9
1976	2.9	7.7	8.8
1977	3.9	7.0	8.0
1978	4.7	6.0	8.5
1979	3.3	5.8	9.2

Sources: *Economic Report of the President, January 1980*, pp. 234, 246, and 247. Wages and hours data for 1979 are from the U.S. Bureau of Labor Statistics.

a. Total hours of work are measured by an index of the hours of all persons in the private business sector. The unemployment rate is the percent of the civilian labor force that is unemployed. Wages are measured by annual averages of compensation per hour in the private business sector.

important element of the total pattern of movement in labor input. Workers put in extra effort during booms and take it easy during slumps. The sum of the declines in productivity in 1974 and 1975 is 3.5 percentage points, a third of the total decline in output of 10.5 percentage points.

The other element of cyclical change in work arrangements that occurs within existing jobs is variation in hours per worker, shown in column 3 of table 2. Both recessions of the 1970s saw pronounced reductions in average hours of work: 2.4 percentage points in 1970–71 and 2.8 points in 1974–75.

Cyclical variations in the employment rate are shown in column 4. Most of these are variations in the number of jobholders and so are outside long-term employment arrangements. The decrease in employment rates characteristic of every recession is only about one-third of the total variation in effective labor input (output); this is a restatement of Okun's Law; namely, a 3 percent variation in real output implies a 1 percentage point change in the unemployment rate.

The last column shows small cyclical variations in the labor force participation rate that represent the other way that labor input can vary during the cycle. The labor force declined by 1 percentage point in the 1970–71 recession and by one-half a point in 1974–75. These variations outside long-term arrangements are quantitatively less important than changes in unemployment.

**Table 2. Components of Fluctuations in Output and Labor Input, 1970–79<sup>a</sup>**

Deviation from trend, in percentage points

<i>Year</i>	<i>Output (1)</i>	<i>Output per hour (2)</i>	<i>Hours per worker (3)</i>	<i>Employment rate (4)</i>	<i>Participation rate (5)</i>
1970	−3.8	−0.6	−1.9	−1.2	−0.1
1971	−0.1	2.0	−0.5	−0.9	−0.9
1972	3.7	2.2	0.5	0.6	−0.3
1973	3.0	0.6	1.3	1.0	0.1
1974	−5.3	−4.3	−0.4	−0.5	0.1
1975	−5.2	0.8	−2.4	−2.9	−0.6
1976	3.6	2.2	0.4	1.1	0.1
1977	2.9	0.6	1.0	0.9	0.5
1978	2.3	−0.8	1.2	1.3	0.8
1979	−0.5	−2.2	1.2	0.4	0.2

Sources: Same as table 1.

a. All data are deviations of annual percentage changes from decade averages. Output is gross domestic product originating in the private business sector at constant prices. Hours are measured by an index of the hours of all persons in the private business sector. The number of workers is civilian employment. The employment rate is the percent of the civilian labor force that is employed. The participation rate is the percent of the population in the civilian labor force.

These findings can be summarized in terms of average percentage point deviations from trend for the two contractions of the 1970s (1970–71 and 1974–75), as shown below.

Change in total output	−7.3
Resulting from decreased work effort	−3.7
Output per hour	−1.1
Hours per worker	−2.6
Resulting from changes in job status	−3.6
Employment rate	−2.8
Participation rate	−0.8

Similar computations could be done for the intervening expansions, but they would be the exact mirror image because the data are detrended. These data on contractions lead to the conclusion that changes in the amount and intensity of effort in existing jobs are an important factor in total cyclical variations in effective labor input. Determination of the intensity of work and hours on the job is an issue just as important as determination of the total number of people at work or in the labor force, though the second issue received the most attention until the recent interest in the theory of long-term employment arrangements. In the next section, I investigate how relevant long-term jobs are in the contemporary American economy.

### The Duration of Jobs in the U.S. Economy

Perhaps the single most distinctive contribution of thinking on macro-labor issues over the past decade has been the investigation of permanent relations between workers and employers.<sup>1</sup> Coexisting, however, has been a body of thought that emphasizes rapid turnover of workers among jobs.<sup>2</sup> High turnover suggests a reasonably fluid market that can be understood in simple market-clearing terms. In particular, the idea that wage payments are installment payments on a long-term financial obligation cannot hold up in a market in which jobs are not of long duration. An obvious starting point in settling the relative importance of the two approaches is an examination of the duration of jobs in the U.S. economy. Oddly enough, I have been unable to find any published studies on this point, so I will present some results derived from a survey by the U.S. Bureau of Labor Statistics on job tenure.<sup>3</sup>

Columns 1 and 2 of table 3 show the distributions of time spent on the current job for workers in two groups, aged thirty to thirty-four and fifty to fifty-four. These numbers come directly from the January 1973 Current Population Survey, which included a question about the date when work-

1. Some of the more important contributions are Martin Neil Baily, "Wages and Employment under Uncertain Demand," *Review of Economic Studies*, vol. 41 (January 1974), pp. 37-50, and "On the Theory of Layoffs and Unemployment," *Econometrica*, vol. 45 (July 1977), pp. 1043-63; Arthur M. Okun, "Inflation: Its Mechanics and Welfare Costs," *BPEA*, 2:1975, pp. 351-90; Costas Azariadis, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, vol. 83 (December 1975), pp. 1183-1202; Martin Feldstein, "Temporary Layoffs in the Theory of Unemployment," *Journal of Political Economy*, vol. 84 (October 1976), pp. 937-57; and Robert J. Barro, "Long-Term Contracting, Sticky Prices, and Monetary Policy," *Journal of Monetary Economics*, vol. 3 (July 1977), pp. 305-16. See also Robert E. Hall and David M. Lilien, "Efficient Wage Bargains under Uncertain Supply and Demand," *American Economic Review*, vol. 69 (December 1979), pp. 868-79.

2. Robert E. Hall, "Why Is the Unemployment Rate So High at Full Employment?" *BPEA*, 3:1970, pp. 369-402, and "Turnover in the Labor Force," *BPEA*, 3:1972, pp. 709-56; Charles C. Holt and others, *The Unemployment-Inflation Dilemma: A Manpower Solution* (Urban Institute, 1971); and Stephen T. Marston, "Employment Instability and High Unemployment Rates," *BPEA*, 1:1976, pp. 169-203.

3. Martin Neil Baily mentions some related findings on job tenure in his "Contract Theory and the Moderation of Inflation by Recession and by Controls," *BPEA*, 3:1976, pp. 585-633.

Table 3. Duration of Jobs among Workers, January 1973

Percent

Duration of job (years)	Workers whose jobs have lasted this long		Jobs that last this long	Workers whose jobs are expected to last this long <sup>a</sup>
	Aged 30-34 (1)	Aged 50-54 (2)	Aged 30-34 <sup>b</sup> (3)	Aged 30-34 <sup>b</sup> (4)
½ or less	15.2	6.5	38.7	3.8
½-1	7.9	3.9	22.6	4.7
1-2	13.3	6.1	9.8	4.7
2-3	9.5	5.4	7.1	4.7
3-5	17.4	11.8	4.6	5.3
5-10	25.0	18.0	8.5	17.3
10-15	10.1	13.3	2.4	8.2
15-20	1.7	10.7	1.1	5.2
20-25	0.1	10.6	0.5	3.8
25-30	0.0	8.4	1.4	10.9
30 or more	0.0	5.4	3.4	31.0

Source: Derived from Bureau of Labor Statistics, *Job Tenure of Workers, January 1973*, Special Labor Force Report 172 (Government Printing Office, 1975), table A, p. A-8. Figures are rounded. See the appendix for details on the derivation.

a. In a population in statistical equilibrium with job durations characteristic of persons aged 30 to 34.

b. This refers to the age of workers at beginning of jobs.

ers began their current jobs. From these observed distributions and others like them for different age groups I computed two other ways of looking at the lengths of jobs.<sup>4</sup> Column 3 of table 3 answers the question: what is the probability that a worker sampled at random from those who have just started new jobs will stay on that job for a particular length of time? The worker considered here is thirty to thirty-four years old. From the results of the computation, it is clear that most jobs are brief. About one-half of all jobs last less than nine months, and three-quarters last less than three years. A few jobs have a long duration—about 9 percent exceed ten years. The average length of a job is just under four years.<sup>5</sup>

From this distribution, one infers that jobs are typically brief just as

4. In this discussion, the reader will note a complete parallel with the literature on the duration of unemployment. See Hyman B. Kaits, "Analyzing the Length of Spells of Unemployment," *Monthly Labor Review*, vol. 93 (November 1970), pp. 11-20; and Kim B. Clark and Lawrence H. Summers, "Labor Market Dynamics and Unemployment: A Reconsideration," *BPEA*, 1:1979, pp. 13-60.

5. Clark and Summers infer the average length of jobs by a completely different procedure based on the frequency of job changes in panel surveys. They conclude that the average is quite a bit lower than the figure reported here. Part of the discrepancy can be explained by a bias toward overstatement of turnover in surveys through random response errors; the rest will require further investigation.

one infers from the corresponding distribution of the length of completed spells of unemployment that they, too, are very brief. However, the brevity of a typical job does not establish that long-term relations between workers and employers are quantitatively unimportant. Exactly because they are brief, short jobs contribute very little to the total volume of work in the economy. The fact that a great many jobs last only a few months does not mean that a large fraction of workers will be employed at brief jobs at any one time. The distribution across workers is obtained from the distribution of job lengths by reweighting by the length of job. The resulting distribution of workers by job length is given in column 4 of table 3. The relative importance of very long jobs then becomes apparent. Half of all workers are in jobs that will last fifteen years or more.<sup>6</sup> About 6 percent of the labor force holds jobs that last nine months or less.

From the perspective of the debates about wage and employment determination, the facts in table 3 can be stated in two relevant ways. First, at any one moment, the majority of workers have not changed jobs in the previous few years. Among workers aged thirty to thirty-four, 54 percent have been employed at their current jobs for three years or more (sum of the last seven figures in column 1 of table 3), and whatever adjustments have been made recently in their earnings and levels of effort have occurred within existing employment arrangements and not as part of initial negotiations. The corresponding figure for workers aged fifty to fifty-four is 78 percent. In any given year, only 23 percent of the younger part of the labor force and only 10 percent of workers aged fifty to fifty-four change jobs. Second, most workers and employers can look forward to continuing relations for quite a few years. Half of all work is done in the course of jobs lasting fifteen years or more (the corresponding figure for men is an astonishing twenty-five years). Any adjustments made currently are likely to be influenced strongly by the prospect of the match continuing for quite a few more years.<sup>7</sup>

6. The distribution used in these computations is from column 3 and so refers to workers who are aged thirty to thirty-four when they start their jobs. In the economy as a whole, of course, there are some workers of all ages who have just begun jobs. Those who are under age thirty typically hold jobs for shorter periods than those shown in column 3 and those who are over age thirty-four for longer periods. These two influences should very nearly cancel, as the median age of recent job starters (those who have taken new jobs in the past six months) is twenty-five years.

7. The empirical relevance of the literature on career labor markets is even greater than many of its contributors may have realized. See Peter B. Doeringer and Michael J. Piore, *Internal Labor Markets and Manpower Analysis* (Heath, 1971).

The data on job durations also show that job turnover is a distinct feature of the labor market, even though it affects only a small number of persons in the labor force. The unemployment resulting from the turnover comes almost entirely from the minority of the labor force that has not settled into jobs lasting three years or more (some temporary layoffs among permanent jobholders contribute to unemployment, as spells of thirty days or less are not counted as job breaks in the data on tenure). The story about short jobs and frequent unemployment told in my 1970 paper is fully supported by these new data on the duration of jobs.<sup>8</sup>

### **Long-Term Employment and the Meaning of Wage Payments**

Recent thinking about relations between workers and employers has emphasized two basic economic motives. First, employers should provide a steady stream of income to workers and let profits absorb temporary fluctuations in demand (*income smoothing*). Second, the marginal value of work should equal the marginal value of time or, more precisely, the marginal rate of transformation between time and goods should equal the marginal rate of substitution between the two (*efficiency*). Some authors have focused on just one or the other, while a few more recent papers have studied the two motives together.

Martin Baily's pioneering paper argued the case for income smoothing persuasively and with considerable generality.<sup>9</sup> Under the reasonable assumptions that firms can borrow and lend and deal with uncertainty more effectively than can individual workers, it makes good economic sense for firms to be financial intermediaries for their employees, spreading total compensation over the duration of the labor contract in a smooth, predictable way. The point is clearest under the following sharply delineated conditions: jobs last, say, ten years, and neither employer nor worker ever breaks the employment contract before it expires. All that matters to the firm is the present value of the total amount of compensation to be paid to the worker. One possible way to schedule compensation would be to pay it in a lump sum at the beginning of the job. Then the worker would be responsible for spreading it over the ten-year span of work to finance a stream of consumption. Another would be to pay it at

8. Hall, "Why Is the Unemployment Rate So High?"

9. Baily, "Wages and Employment under Uncertain Demand."

the conclusion of the ten years, which would require the worker to borrow to finance consumption in the interim. It is more convenient for both employer and worker if compensation is a stream over the course of the job, as this arrangement limits the worker's need to make large asset transactions. Further, issues of reliability and trust make it desirable that neither party be heavily in debt to the other. Where compensation is paid as a stream, it should be viewed as an installment payment on the firm's long-term obligation to the worker.

Home mortgages provide an interesting analogy to long-term labor contracts under the extreme assumptions of inviolable long-term employment commitments. It is generally convenient to the home owner to spread payments over a long period rather than making a single payment either at the time of purchase or at some later date. The choice of schedule for the payment is almost purely a question of the convenience for the home owner—the bank cares only about the present value of the payments. The terms of the mortgage are settled when the contract is signed; they may be contingent on outside variables, as in a variable rate mortgage, but there is no good reason for them to respond to the current state of the housing market. One way of putting the basic argument for sticky wages under labor contracts is that there is equally little reason to expect the current flow of compensation under a wage contract to reflect the current state of the labor market. To see what is happening in today's housing market, one looks at current prices being paid for houses, not at average mortgage payments of home owners. Similarly, to see what is happening today in the labor market, one should look at the implicit asset prices of labor contracts recently negotiated, not at the average rate of compensation paid to all workers. This point is familiar in interpreting wages in organized industries with three-year formal contracts, but is much more thoroughgoing in an economy in which most workers have jobs that will last more than fifteen years. Even in organized industries, it would be farfetched to suggest that everything starts again each time a new formal contract is negotiated.

The fact that most workers remain on their jobs for long periods does not itself establish that long-term employment arrangements differ from those in a completely open spot market. One would not want to argue that the stockholders in IBM had special long-term associations with their individual shares just because they typically own them for fifteen years.<sup>10</sup>

10. Martin Feldstein pointed out this false analogy.

There must be some economic glue in the form of specific human capital binding workers to jobs for the long-term arrangement to have the implications discussed here. Without the glue, credible threats of workers to quit and employers to lay off workers would push current compensation to the point where it reflected the current marginal product of labor and current marginal value of workers' time. The firm's ability to act as a financial intermediary to smooth compensation may provide some of the glue if independent intermediaries do not function in the labor market. I assume for the remainder of this section that the glue is sufficiently strong that employment arrangements are effectively full bilateral commitments for quite a few years.

At the simplest level, the schedule of wage payments for the duration of an unbreakable employment contract is a matter of indifference. Evidence recently assembled by James Medoff suggests that large U.S. corporations tend to use contracts with rather extreme back-loading—older workers are generally paid more than younger ones until retirement age even though productivity tends to decline in later years.<sup>11</sup> For a considerable period after the beginning of the typical employment arrangement, the firm accumulates a growing debt to the worker. Later in the worker's career, this debt is partly drawn down by higher wages. The rest is paid off in the form of retirement benefits. Edward Lazear argues that mandatory retirement is an essential feature of employment contracts with this kind of back-loading.<sup>12</sup> Various explanations of back-loading of employment contracts have been offered, based on problems in supervising workers, reducing incentives to quit, and the like. What is important for macroeconomics, though, is simply the existence of long-term employment contracts and the unresponsiveness of wage payments under them to current economic conditions.

The desirability of efficient labor contracts has become a prominent

11. James L. Medoff, "The Earnings Function: A Glimpse Inside the Black Box," Discussion Paper 594 (Harvard Institute of Economic Research, December 1977); and James L. Medoff and Katharine G. Abraham, "Experience, Performance, and Earnings," forthcoming in *Quarterly Journal of Economics*, "Are Those Paid More Really More Productive? The Case of Experience," forthcoming in *Journal of Human Resources*, and "Involuntary Terminations under Explicit, Implicit, and No Employment Contracts" (Harvard University, 1980).

12. Edward P. Lazear, "Why Is There Mandatory Retirement?" *Journal of Political Economy*, vol. 87 (December 1979), pp. 1261–84.

feature of the more recent literature.<sup>13</sup> Every labor contract must provide a rule by which the level of effort of the worker is determined. In the earliest models the choice was binary—the worker was either at the job full time or not at all. A more elaborate contract can specify variations in daily or weekly hours. In all cases, the natural economic assumption is that the rule comes as close as it can to an efficient outcome. Under the contract, the level of employment has the property that no other level could make both the firm and its workers better off. Another way to put it is: employment determination is a bargaining problem between them. An efficient level of employment is a point on the contract curve of that bargaining problem, and sensible bargainers ought to be on the contract curve.<sup>14</sup>

From the private points of view of the employer and the worker participating in a labor contract, efficiency requires that the value of the gains to the employer from additional employment equal the value of whatever the worker has to give up in order to supply the additional work. What the employer gains is the marginal revenue product of labor. What the worker gives up is more complicated. A worker may have to reduce hours on a second job, in which case it is easy to estimate a dollar value for what is given up. In most cases, though, the worker sacrifices nonmarket activities, and it is more difficult for the analyst to put a value on them. The duration of an increase in time spent at work is an important consideration—workers are probably much more willing to work long hours for a few weeks or months than for several years.

The same considerations apply when the firm contemplates a reduction in employment—the forgone revenue should equal the value to the worker

13. An early example in which efficiency is the prime goal of a labor contract is Feldstein, "Temporary Layoffs." A full treatment of the conditions for efficiency appears in Baily, "On the Theory of Layoffs and Unemployment." My own paper with David Lilien also emphasizes efficiency; see our "Efficient Wage Bargains." An important paper that tries to find the optimal blend of smoothing and efficiency is Guillermo A. Calvo and Edmund S. Phelps, "Employment Contingent Wage Contracts," in Karl Brunner and Allan H. Meltzer, eds., *Stabilization of the Domestic and International Economy*, Carnegie-Rochester Conference Series on Public Policy, vol. 5 (Amsterdam: North-Holland, 1977), pp. 160–68.

14. The point that the bargain involves an explicit employment dimension as well as compensation is made very effectively by Wassily Leontief, "The Pure Theory of the Guaranteed Annual Wage Contract," *Journal of Political Economy*, vol. 54 (February 1946), pp. 76–79.

of spending more time in other activities. Again, it seems likely that workers can make much better use of temporary increases in time at their disposal than permanent ones: a brief spell away from work is a good time to go on vacation or to repair the house, but these opportunities are exhausted in a few weeks or months. The evidence on the labor supply of adult males seems to indicate that small permanent variations in hours of work cause very large variations in the marginal value of time.<sup>15</sup> Low compensated wage elasticities of labor supply mean that workers feel they give up a lot if they work more than normal full time and derive very little personal benefit from working less than full time. Efficient contracts for them would promise that they work almost exactly full time on the average.

Private considerations of efficiency may not coincide with social considerations if the tax system drives wedges into private calculations. Taxation of business and wage income reduces the private return to employment below the social return and pushes private labor contracts to levels of employment that are socially inefficiently low. Another tax consideration has been prominent in the discussion of labor contracts: workers are paid unemployment compensation during periods of layoff, and much of this is not recovered from their own employer through insurance premiums, but is paid from taxes levied on employers in general. This tax wedge pushes employment contracts in the direction of accommodating decreases in demand with temporary periods of nonwork. Workers perceive these temporary layoffs as privately valuable because of the unemployment compensation they receive, but there is no corresponding social value to the resulting unemployment. Feldstein gives a complete mathematical statement of the private efficiency condition for setting employment in the face of an unemployment insurance system that pays benefits during periods of temporary layoffs.<sup>16</sup> He argues persuasively that such a system raises the natural rate of unemployment, but the effect cannot be large because temporary layoffs are a small part of full-employment unemployment.

In one form of employment contract, the employment rule is completely separate from the compensation rule. Compensation is a lump

15. Robert E. Hall, "Wages, Income, and Hours of Work in the U.S. Labor Force," in Glen G. Cain and Harold W. Watts, eds., *Income Maintenance and Labor Supply: Econometric Studies* (Rand McNally, 1973), especially p. 153.

16. Feldstein, "Temporary Layoffs."

sum paid by the employer to the worker in exchange for the agreement to supply effort according to an implicit employment agreement. The employer informs the worker each week about that week's efficient level of effort, or, for more responsible workers, the decision may be joint or even unilateral by the worker. In unusually busy weeks, longer hours are set and workers may also accomplish more in each hour. In slack weeks, hours are set at lower levels and the intensity of work may fall as well. The general flavor of the arrangement is that employees work harder if there is more work to do. I think this is a reasonable summary of the employment bargain for many salaried white-collar workers (who now constitute about half the total labor force). Employers have the right to request intense effort for a few weeks or months, but not permanently. Periods of extraordinary effort must be counterbalanced by restful periods. For salaried workers, arrangements of this kind develop by custom and are rarely spelled out in formal contracts. The absence of a relationship between weekly effort and weekly compensation is virtually the definition of a salaried job.

Another form of employment contract is studied in my paper with David Lilien.<sup>17</sup> The problem again is to create a workable mechanism for determining employment in a way that respects the value of labor's time. We ask if labor can set up economic inducements for management to adopt an efficient level of employment; we have in mind explicit bargaining between blue-collar industrial workers and management in a collective bargaining setting. The union does not trust management with the unfettered unilateral power to set employment and hours because it fears it will be asked to work systematically too hard. Nonetheless, it recognizes the mutual benefits of working harder when there is more work to do. The contract we suggest requires management to compensate labor for the marginal opportunity cost of the hours that labor is required to supply. Under the right compensation formula, management fully internalizes labor's interests, and so makes an efficient decision unilaterally.

Contracts of this kind have the attractive feature of accommodating large shifts in the demand for output without renegotiation. In a situation in which management is likely to have better information about the current state of demand for products than does labor, it is sensible for management to make a unilateral decision about employment. The alternative is to make the employment contract contingent on some measurement of

17. Hall and Lilien, "Efficient Wage Bargains."

demand, but such contingencies have problems of moral hazard and verifiability that have been pointed out by a number of authors.<sup>18</sup> To induce the efficient response of employment to a shift in demand, the compensation formula must embody labor's views about the opportunity costs of brief variations in work schedules in contrast to long-term variations. For example, compensation could drop sharply for a brief layoff because unemployment compensation was available at no cost to either employer or worker or because workers can make good private use of a few weeks off the job. On the other hand, firms would be given much less opportunity to lower compensation costs through permanent reductions in hours of work. In periods of strong demand, extra hours could be obtained temporarily at modest increases in compensation, but again, permanent increases would be penalized. If the long-run supply of labor is as inelastic as the evidence suggests, the efficient contract would effectively prohibit permanent changes in hours, so the contract might simply state this as a rule rather than imposing severe economic penalties. In fact, collective bargaining agreements typically contain many pages describing the provisions for varying hours, putting workers on temporary layoff, promoting and demoting workers (thereby changing their compensation), and so forth. A key notion is that management's unilateral role in setting the volume of work is carried out within carefully specified provisions of a contract.

I argue that the collective bargaining agreements found in the contemporary United States fit in quite well with this prescription. Financial disincentives for excessive weekly hours of work (in the form of overtime premiums) do protect workers against being told to work hard every week. Limitations on the number of consecutive weeks of overtime have the same effect. On the other hand, many contract provisions limit the cost savings from reducing labor input below normal. In a number of major industries, firms are committed to keeping workers' incomes at normal levels, even during temporary layoffs, through supplemental unemployment benefits. Another widespread provision of collective bargaining agreements requires concentration of layoffs among the lowest-paid workers, again limiting the cost savings from layoffs. As a general matter,

18. See, for example, Calvo and Phelps, "Employment Contingent Wage Contracts"; Barro, "Long-Term Contracting"; and Stanley Fischer, "Long-Term Contracts, Rational Expectations, and the Optimal Money Supply Rule," *Journal of Political Economy*, vol. 85 (February 1977), pp. 191-205.

agreements contain a great variety of incentives and rules for stabilizing employment to respect the value of workers' time.

A limitation of contracts of this kind is their inability to make employment respond efficiently to shifts in the opportunity costs of labor. A contract cannot be written that will make management internalize unexpected shifts in the value of labor's time without introducing contingencies. One contingency—changes in the cost of living—is present in many collective bargaining agreements, but I am not aware of any other formal contingencies. Again, there are important practical obstacles to contingent contracts. In our joint paper, Lilien and I suggest that periodic renegotiation of agreements occurs for the specific purpose of taking account of inefficiencies that enter on the supply side.

Contracts that grant management the unilateral right to set the volume of work will create changes that are viewed by the workers themselves and by observers as involuntary. Workers do not consider the current values of the relevant variables and then decide how much to work. They will simply do what they are told by management, relying on the accuracy of management's computations of the efficient, profit-maximizing level of employment. Fluctuations in employment could be efficient even though they are involuntary—their involuntary nature is not by itself a conclusive case for inefficiency.<sup>19</sup>

### **Job Termination**

Long-term employment contracts cannot provide complete stability of employment, nor would it usually be desirable for them to do so. Forward commitments by workers are almost invariably unenforceable legally. Commitments to provide employment in future years are enforceable, but

19. This distinction is not widely appreciated. The presidential addresses of James Tobin, Franco Modigliani, and Robert Solow before the American Economic Association point to the involuntary nature of layoffs as evidence against the hypothesis that fluctuations in employment are efficient. See James Tobin, "Inflation and Unemployment," *American Economic Review*, vol. 62 (March 1972), pp. 1–18; Franco Modigliani, "The Monetarist Controversy or, Should We Forsake Stabilization Policies?" *American Economic Review*, vol. 67 (March 1977), pp. 1–19; and Robert M. Solow, "On Theories of Unemployment," *American Economic Review*, vol. 70 (March 1980), pp. 1–11. All three have many other arguments against the hypothesis as well. Solow has written to me in defense, "An old non-com like me knows perfectly well that efficient decisions can be handed down against the will of the decidees—and inefficient ones too."

indefinite employment may not be efficient. If there is a permanent fall in the demand for a firm's products, it is probably efficient to reduce the total labor force and have the remaining workers put in full-time hours rather than maintaining the labor force at its earlier level with shorter hours. Efficient contracts need to provide a way to terminate jobs as well as a way to vary total labor input.<sup>20</sup>

For jobs with an important component of specific human capital, the open market cannot support efficient job termination through unilateral decisions made by workers and employers: if wages include the full return to the specific capital, workers have the appropriate incentives about quitting, but employers have no incentive to preserve the specific capital by holding back on layoffs; if employers receive the return, then workers are too likely to quit. One way to achieve efficiency is to require the worker and the employer to compensate the other for breaking the job match.<sup>21</sup> A more realistic approach is for rules to evolve that limit layoffs and quits and permit them to occur only when they are efficient. Layoffs should be allowed only under conditions of genuine permanent reduction in demand and not just when it is privately beneficial to the firm. Further, layoffs ought to be concentrated among junior workers whose specific capital is smaller and whose costs of finding new work are probably lower. For workers, quits should occur only for good personal reasons—a distinctly better job elsewhere, an opportunity to acquire additional education, and the like—not merely because of a modest increase in wages in a different job. The practices and implicit contractual provisions supporting these limitations are likely to assign the employer the responsibility for terminations made efficient by declines in demand and the worker the responsibility for quits made efficient by superior alternatives elsewhere. Thus layoffs and quits have distinct economic meanings.<sup>22</sup>

Economists have only just begun to examine the issues in the efficient movement of workers among firms. The obstacles to efficiency here are

20. In the abstract, these are really the same issue because permanently working zero hours at a job is about the same as not having the job.

21. Gary S. Becker, Elisabeth M. Landes, and Robert T. Michael, "An Economic Analysis of Marital Instability," *Journal of Political Economy*, vol. 85 (December 1977), pp. 1141–87, especially p. 1145 and note 4.

22. In Becker, Landes, and Michael's arrangement based on individual bargaining between workers and employers, there is no meaningful distinction between layoffs and quits. In practice, the classification is often ambiguous; the U.S. Bureau of Labor Statistics tends to classify all ambiguous job terminations as layoffs.

much greater than in the case of a permanent match between worker and employer. When a worker is about to move from one firm to another, there are three participants in the implicit bargain: the worker, the current employer, and the future employer. Institutions like unilateral permanent layoffs that work reasonably well in normal times when the departing worker has a good chance of lining up a new job quickly may not be efficient in times of recession. The failure of macro efficiency over the business cycle may be attributable in part to the inability to create the rather complex institutions necessary to deal with movements of workers among firms.

### **Short-Run Sensitivity of Marginal Value of Time to Hours of Work**

In order for the theory of efficient employment determination to explain fluctuations in hours of work and other dimensions of labor input, the marginal value of time cannot be too sensitive to the volume of work. The principle that employees work harder when there is more work to do fails if the marginal valuation of the additional work is enormous. The issue here is the short-run labor supply function of the typical worker in the following sense: the labor supply function shows the individual's offer of work to the labor market under the assumption that the market will absorb any amount of that work at the prevailing wage. Under these conditions, the labor supply schedule traces out exactly the worker's marginal valuation of time. Fluctuations in the level of work will occur, then, if the short-run labor supply schedule is somewhat responsive to the wage. However, a central feature of the theory of employment determination under long-term arrangements is the very different institutional procedure for setting employment compared to an open labor market. Few workers carry out an explicit labor supply calculation every month or year. Instead, they just work as much as they are told. But the labor supply issue arises just as importantly in the following way: when an employer unilaterally imposes an increase in hours, what is the dollar value of the monetary and psychic costs to the worker?

One body of research with a claim to answering this question is standard econometric studies of labor supply. But these studies give seriously misleading answers if applied to workers under long-term employment arrangements. They universally assume workers are free to vary their hours unilaterally and receive the same wage for each extra hour. For

example, if some workers in a sample hold jobs with heavily back-loaded compensation arrangements, the assumptions of standard labor supply studies would let them raise hours in the later years of high compensation. In fact, they do not have this opportunity, so the results understate the amount people would vary their hours if presented with a genuine opportunity to do so in the short run. Econometric work on labor supply has been criticized by institutional labor economists on exactly these grounds since the earliest studies. Recent thinking on long-term employment arrangements clearly supports this criticism.

Another body of research is based on the negative income tax experiments. Workers were given substantial cash grants at the same time that their wages were taxed at 30, 50, or 70 percent. Again, one could argue that their labor supply responses were attenuated by contractual provisions in which employers had the unilateral right to set hours of work. But this point has much less force in a low-income population, where turnover is high and long-term employment arrangements are rare. Labor supply responses to the taxes and transfers in the experiments have been strong—about 14 percent reduction in hours for men and 75 percent for women.<sup>23</sup> If all of this is attributed to the temporary reduction in after-tax wages (on the ground that temporary increases in income should have little effect on labor supply), elasticities of short-run labor supply of 0.2 to 2.0 emerge. But investigators have reached differing conclusions about the strength of the effect of the income transfer. The most that can be said at this point is that hours of work do respond sharply to temporary changes in economic circumstances.

Still another approach is direct measurement of the opportunity cost of work. Martin Feldstein estimated that forgone unemployment benefits are 40 to 50 percent of regular compensation for the typical eligible worker.<sup>24</sup> Although unemployment benefits are paid in most states only for full weeks of unemployment, they effectively create a marginal op-

23. Robert E. Hall, "Effects of the Experimental Negative Income Tax on Labor Supply," in Joseph A. Pechman and P. Michael Timpane, eds., *Work Incentives and Income Guarantees: The New Jersey Negative Income Tax Experiment* (Brookings Institution, 1975), pp. 126 and 128, and the comments by Zvi Griliches following the chapter, p. 147.

24. Martin Feldstein, "Unemployment Compensation: Adverse Incentives and Distributional Anomalies," *National Tax Journal*, vol. 27 (June 1974), pp. 231–44, and "Theory of Temporary Layoffs." These computations include the offsetting effects of experience rating, which result in making the employer pay part of the benefits.

portunity cost of raising year-long weeks of work, up to the point of no temporary layoffs throughout the year. With respect to the value of activities that must be given up in order to work more, estimates of foregone job search reported by Robert Gordon are small.<sup>25</sup> So far as I am aware, nothing is known about temporary jobs taken by workers on temporary layoff. And finally, no direct evidence is available about the value of nonmarket activities of workers undergoing temporary reductions in market work.

In summary, there is relatively little good evidence on workers' valuation of short-run variations in hours of work. Except for the purely financial rewards from unemployment compensation, analysts must rely on introspection to decide if the typical worker makes good use of temporary spells away from his job.

### **Wage Payments and Wage Inflation under Long-Term Contracts**

Under a long-term contract the current rate of hourly compensation has little relation to the current marginal product of labor or the current marginal value of time. The separation of wage payments from current economic conditions is most extreme if wages are purely installment payments on long-term obligations. In long-term salary contracts, the current wage has no allocational role. The labor market is effectively an asset market, and the implicit present value of future salary payments has to be observed to measure the price that is clearing the market.

Under the contracts discussed in my paper with Lilien, which we argue are representative of arrangements for blue-collar workers under collective bargaining, marginal compensation is set up to guide employment decisions and does have an allocational function. In principle, information could be derived from the joint variation of hours and compensation under this type of contract.<sup>26</sup> But contract provisions should make average compensation fairly insensitive to the level of employment, even if

25. Robert J. Gordon, "The Welfare Costs of Higher Unemployment," *BPEA*, 1:1973, pp. 133-95.

26. A related idea was developed in my "The Process of Inflation in the Labor Market," *BPEA*, 2:1974, pp. 343-93. There I looked for marginal compensation in excess of average compensation. But a more complicated relation is suggested by my work with Lilien: marginal above average for more hours than normal and marginal below average for fewer hours than normal. See Hall and Lilien, "Efficient Wage Bargains."

marginal compensation is quite sensitive. Again, the current hourly wage is not the quantity guiding the employment decision of the firm.

How does the Phillips curve showing the trade-off between inflation and unemployment fit into the new ideas and facts about long-term employment arrangements? The notion of a stable relation between unemployment and inflation fared poorly in the past decade. Of course, few economists ever said that price changes would fit a simple Phillips curve when the world prices of raw materials gyrated. But even wage inflation had a life of its own relative to unemployment during the decade, and the upward shift of the Phillips curve associated with rising inflationary expectations contributes relatively little toward understanding the movements of wages in the 1970s. The single largest disappointment for the idea that demand affects wage inflation was the persistence of high wage inflation in the year of the deepest postwar recession, 1975.

In an economy in which most compensation is paid under long-term employment arrangements, the insulation of wages from the current state of demand is not too great a surprise. The wage does not clear the market; the employment provisions of contracts do that, if indeed the market is clearing. Wages are installment payments on long-term financial obligations. As William Nordhaus has put it, “. . . contracts are written over what might be called the economic climate rather than the economic weather.”<sup>27</sup> However, the magnitudes of the obligations cannot be set unambiguously in advance, if only because of uncertainty about the future value of the dollar. Adjustments need to take place to accommodate surprises in monetary and fiscal policy, shifts in the relative prices of food and oil, and many other unforeseen developments. Annual rates of wage inflation will reflect whatever changes have been made in long-term obligations, changes that were planned in advance, changes that have occurred in contracts with compensation formulas linked to current employment, and full market-clearing movements among the minority of workers in the open labor markets with short-term employment arrangements. Plainly, such a hodgepodge of sources of wage movements will not have a simple relation to a single measure of the state of demand.

The wage equations fit by Robert J. Gordon reflect the diversity of considerations entering wage determination.<sup>28</sup> The recent behavior of

27. William D. Nordhaus, discussion of Baily, “Contract Theory,” p. 623.

28. Robert J. Gordon, “Can the Inflation of the 1970s be Explained?” *BPEA*, 1:1977, pp. 253–77 and the earlier papers cited there.

consumer prices, the price of domestic value-added, payroll and income taxes, federal intervention through guideposts and controls, and real output all have important roles in addition to a variety of measures of unemployment. Even though Gordon is particularly concerned with careful measurement of the effect of demand, he finds an extremely flat Phillips curve. His results suggest that an increase of 1 percentage point in the unemployment rate sustained for a full year would depress wage inflation by one-quarter to one-half a percentage point. When combined with traditional theories of employment determination in which the marginal product of labor is equated with the current wage, these results give a terribly pessimistic prediction about the ability of the economy to restore full employment after a negative shock, without government intervention.

Because the current rate of compensation is not directly related to employment determination, the Phillips curve is off center stage in an economy in which long-term employment arrangements achieve macro efficiency. In such an economy, the Phillips curve deals with the distribution of income but not with the allocation of resources. Rather than study the observed flow of compensation to workers, economists should be looking at the longer-run terms offered to workers taking new jobs.

But before consigning the Phillips curve to the scrap heap, the failure of the efficiency principle at the macro level should be noted. In the next section of the paper I discuss how implausible it is to assert that the labor market is fully successful in allocating labor efficiently over the business cycle. The longer-run terms offered to new workers may not adjust to clear the labor market, but instead may follow something like the Phillips curve. Since new thinking about long-term employment arrangements has not reached the point of explaining the observed fluctuations in aggregate employment and unemployment, economists should not be too hasty in discarding the old theory in which the nominal wage as determined by the Phillips curve has an important allocational role.

### **The Failure of Macro Efficiency**

Theories of efficient, long-term employment arrangements have made an impressive contribution to understanding the labor market, but they are far from closing the debates about the nature of the business cycle and employment fluctuations. There are three ways in which the evidence

does not sustain the efficiency of cyclical fluctuations in employment and unemployment.

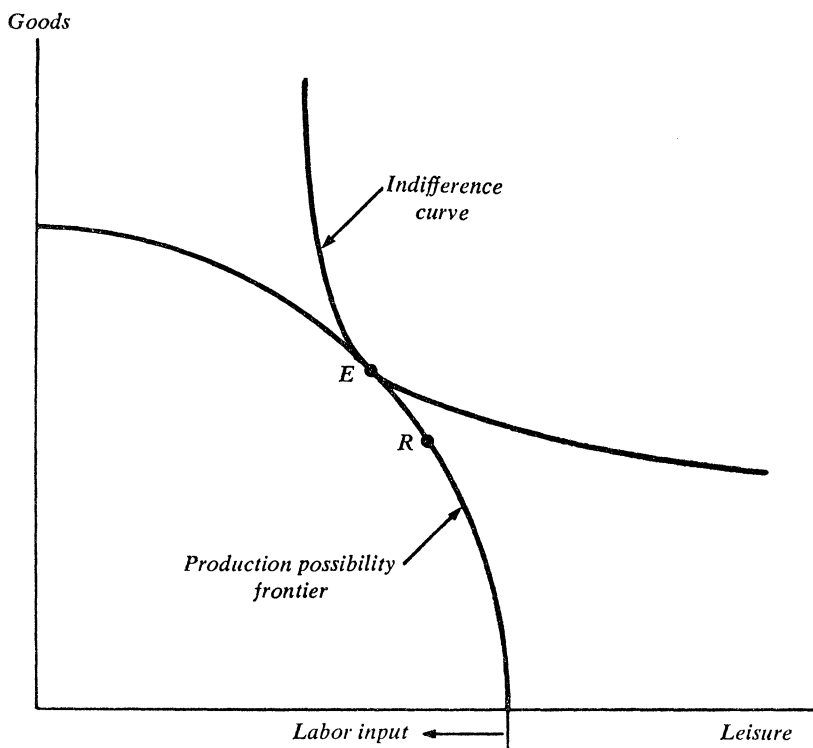
First, the proposition that employment arrangements are efficient explains a lot about fluctuations in employment in individual firms, industries, and sectors; but in the aggregate it is unclear what economic forces can create efficient fluctuations in the amount of work there is to do. Simple ideas of efficiency seem to suggest highly stable employment at the macro level.

Second, the efficient level of employment is a real economic quantity and should be unaffected by purely nominal influences. But ample evidence suggests that changes in the stock of money have a lingering effect on employment.

Third, efficient employment arrangements create chronically positive levels of unemployment, as micro fluctuations will continually cause workers to be put on temporary layoff or to find work in other firms or sectors. However, the data suggest that relatively little of the overall level of unemployment can be explained in this way.

### **Are Fluctuations in Aggregate Employment Efficient?**

Up to this point, I have discussed the problem of efficient employment determination at the level of the firm. Variations in product demand could be local to the firm, industry-wide, or part of an economy-wide contraction. The ideas reviewed seem incapable of explaining fluctuations in a simple aggregate economy. The reason is shown in the diagram below. At full employment, the efficient allocation of time to the production of goods is the point of tangency of the production possibility frontier to the indifference curve, at point *E*. During a recession, the economy moves to a point like *R*, with less labor input and less production of goods. In order to interpret this as a new efficient point, either the production possibility frontier must have twisted counterclockwise (a sudden decline in the productivity of labor) or the indifference curves must have twisted clockwise (a downward shift in labor supply). Neither seems to be a good description of a recession. If a second productive factor is introduced into the analysis, say oil, a temporary increase in its price might have effects similar to those of a decline in productivity. But recessions have occurred many times without oil price increases, and there have not



been any temporary oil price increases yet. Twisting of the indifference curve seems equally implausible—an explanation of recessions based on temporary shifts of labor supply functions (not movements along them) seems completely empty.<sup>29</sup>

29. Recent work by Ben Bernanke investigates explanations of aggregate fluctuations within theories of efficient employment arrangements. His economy has two sectors, durable and nondurable goods. For good economic reasons, the economy undergoes a pause in its efficient rate of accumulation of durable goods. Workers do not shift costlessly to the nondurable sector for the usual reasons of specific human capital and the like. Instead, they work short hours and enjoy some time off the job until the demand for durable goods picks up. All this rests on the idea of efficient employment arrangements in the durable goods sector and an elasticity of the marginal value of time with respect to the level of work that is not too high. Although this seems a fruitful line of research at this point, it would take me too far from the topic of the paper to appraise its empirical success. See Ben S. Bernanke, "Long-Term Commitments, Dynamic Optimization, and Business Cycles" (Ph.D. dissertation, Massachusetts Institute of Technology, 1979).

## Monetary Influences on Employment

The efficient level of employment, at which the value of time equals the marginal product of labor, does not depend on the stock of money or any other nominal quantity; it is exclusively real. This implication is flatly contradicted by a large body of evidence on the positive correlation between money and real output and employment.<sup>30</sup> Moreover, movements of the money stock often precede changes in employment, so it is implausible that money is responding to influences from the real economy.

More than any other evidence, the vulnerability of employment to nominal shocks supports the traditional theory of employment determination and makes the relevance of models based on efficiency questionable. If, as the traditional theory asserts, the nominal wage is predetermined and employers are free to treat it as the marginal cost of raising or lowering employment, explaining the response of employment to movements of the money stock is a simple textbook exercise. A decade of research by macroeconomists who are dissatisfied with the absence of a microeconomic foundation for the traditional theory has not produced a workable alternative based on realistic assumptions.<sup>31</sup>

## Unemployment

Under efficient employment arrangements with unilateral decision-making by employers, unemployment ought to have two components: workers on temporary layoff who will be recalled fairly soon, and workers who have been discharged permanently because they belong in other jobs. A more fully developed theory of efficient labor markets would con-

30. See, for example, Robert J. Barro, "Unanticipated Money Growth and Unemployment in the United States," *American Economic Review*, vol. 67 (March 1977), pp. 101–15. Whether or not Barro is successful in isolating the influence of unanticipated money on employment is not important here. What we get from his work is the unambiguous conclusion that lagged money predicts current employment.

31. Robert Lucas' rigorous demonstration of monetary influences on real variables in a simple model with limitations on the diffusion of information has not so far been transplanted to a more realistic economy where information about most economic developments is widely available within a month. See Robert E. Lucas, Jr., "Expectations and the Neutrality of Money," *Journal of Economic Theory*, vol. 4 (April 1972), pp. 103–24.

**Table 4. Unemployment by Reason, 1977**

<i>Reason for unemployment<sup>a</sup></i>	<i>Number of unemployed (thousands)</i>	<i>Percent of total unemployment</i>
On layoff	853	12
Temporary	234	3
Indefinite	620	9
Lost job	2,250	33
Left job	889	13
Wanted temporary work	924	13
Left school	469	7
Other	1,470	21
Total	6,855	100

Source: Bureau of Labor Statistics, unpublished tabulations from the Current Population Survey.

a. The categories "wanted temporary work," "left school," and "other" are alternatives to the published breakdown into entrants and reentrants. "Wanted temporary work" is one of the possible answers to the question on the survey on why an individual became unemployed. It is unclear how a worker is categorized who becomes unemployed by losing or leaving a temporary job. Presumably most people in this category have entered the labor force to look for temporary work. Figures are rounded.

sider a small amount of unemployment from people looking for their first jobs or for new jobs after a period outside the labor force and also would consider job quitters, some of whom would become unemployed. The actual composition and volume of unemployment in 1977 is shown in table 4. The overall level of unemployment was 7.0 percent, a little above the average for the decade of 6.2 percent and somewhat above most estimates of the natural or equilibrium rate for the 1970s.

The first row of the table shows the number of workers who are classified as "on layoff," meaning that they still have jobs but are on furlough. This group, which figures prominently in most discussions of efficient employment arrangements, especially in an economy with generous unemployment compensation, actually accounts for only 12 percent of the unemployed. Of these, about a quarter are expecting recall within thirty days (the temporary category) and the other three-quarters expect to return to work eventually but not within thirty days (the indefinite category). In a nonrecessionary year like 1977, temporary layoffs from permanent jobs are not a major component of total unemployment. The fourth row of table 4 shows that job losers form a much larger group; these are workers who are looking for new jobs and do not have any reason to expect to return to jobs that have ended. Many people in this category have been laid off in the conventional sense that their employers have unexpectedly told them to stop work. But it also includes large

numbers whose jobs were explicitly temporary and have simply come to an end. The job losers fit the standard picture of the unemployed: they have lost jobs and are looking for new ones.

The next group in table 4 is the unemployed quitters, who make up only 13 percent of total unemployment. Quitting is a common way for people to leave jobs, but generally they do not become unemployed subsequently. Instead, they move directly to new jobs or out of the labor force. Another 13 percent of the unemployed associate their joblessness with an interest in temporary work, though the question on the unemployment survey that elicits this response is so vague that interpretation is difficult. Next, people just out of school make up 7 percent of the unemployed. This is a dramatic illustration of one of the central themes of recent findings on unemployment: net flows, like the movement from school to work, are tiny compared to the gross flows from job to unemployment to job or into and out of the labor force. Finally, the category "other" comprises primarily people who have been out of the labor force, usually for no more than a few weeks or months.

Nothing in table 4 refutes the idea that efficient employment arrangements lie behind the rather high level of unemployment that is characteristic of the U.S. labor market in many years, but it would also be hard to say that they support the hypothesis. The type of unemployment that is most closely identified with the hypothesis of efficient use of workers' time, namely temporary layoffs, is a very small part of total unemployment. A more detailed analysis of the other components of unemployment yields the following: a substantial amount of unemployment comes not from workers who occasionally spend a few weeks away from their jobs because there is no work to do, but from people who occasionally spend a few months working but are looking for work or are out of the labor force most of the time. Although this assertion was made by a number of earlier authors, Kim Clark and Lawrence Summers found the most dramatic ways of expressing it.<sup>32</sup> Only 28 percent of total unemployment in 1974 was contributed by spells of two months or less that ended by finding work. Almost half (47 percent) of all unemployment comes from spells that ultimately end in withdrawal from the labor force. Almost half (45 percent) of all unemployment comes from spells lasting five months or more. Clark and Summers also document the concentra-

32. Clark and Summers, "Labor Market Dynamics and Unemployment."

**Table 5. Unemployment in Tight and Slack Markets, by Reason, 1974 and 1975**

Percent of labor force

<i>Reason for unemployment</i>	<i>Tight market, 1974</i>	<i>Slack market, 1975</i>
On layoff	0.8	1.8
Lost job	1.6	2.9
Left job	0.8	0.9
Wanted temporary work	0.9	1.1
Left school	0.4	0.5
Other	1.1	1.3
Total	5.6	8.5

Source: Same as table 4.

tion of unemployment in a minority of the labor force. In 1974, only 15 percent of the civilian labor force out of school had any unemployment at all. The average amount of unemployment among those with at least one week was three and a half months. The 2.4 percent of the labor force with more than six months of unemployment accounted for 41.8 percent of all unemployment.

The facts about episodic employment reveal how incomplete research has been to date on the nature of employment. The tendency to think of workers as spending most of the year at work is appropriate for most of the labor force; but for the type of worker who dominates unemployment, that tendency is quite inappropriate. No research has established that episodic work is an inefficient use of time; rather, the subject has been almost completely overlooked by research. For now, all we can say is that it is an important phenomenon of the labor market that we do not understand.

Theories of long-term employment arrangements are somewhat more successful in explaining the cyclical behavior of unemployment than they are in explaining the level of unemployment. Table 5 shows the changes in the composition and the level of unemployment brought about by a recession. The table breaks down unemployment into the same categories as in table 4 for a year of tight labor markets, 1974, and a year of deep recession, 1975. Of the increase in total unemployment of 2.9 percentage points, 1.0 occurred for workers on layoff and 1.3 percentage points for those who lost jobs. The other 0.6 point is spread among those who left jobs, wanted temporary work, left school, or were in the "other" category. About 35 percent of the increase in unemployment during a

contraction occurs among workers who have not definitely lost their jobs, which is in accordance with the ideas of long-term employment arrangements. However, the fraction of unemployed workers who are on layoff declines dramatically by the beginning of the expansion, long before total unemployment returns to normal.

## **Conclusions**

It is clear why employees in one firm work harder when there is more work to do in that firm, but not why there is more total work to do in the aggregate economy in some years than in others. Institutional arrangements in the labor market like temporary and permanent layoffs and unresponsive wages make good economic sense for individual firms dealing with their own fluctuations in demand, but it is not known why they sometimes operate in unison to depress employment throughout the economy.

The greatest recent progress in understanding the labor market comes from the study of long-term employment arrangements. There is no point any longer in pretending that the labor market is an auction market cleared by the observed average hourly wage. In an extreme case, wages are just installment payments on a long-term debt and reveal essentially nothing about the current state of the market. Because wages are not necessarily the appropriate guide to employment decisions under long-term employment arrangements, analysis has turned to employment determination as an important issue in its own right. But the simple microeconomic idea of efficient levels of employment does not seem to be able to explain the large observed fluctuations in the level of employment over the business cycle. The traditional idea of sticky nominal wages and unilateral profit maximization by employers has hardly been overturned by the new ideas. It seems safe to predict that a good deal of additional effort will be expended in the forthcoming decade in trying to improve microeconomic understanding of cyclical variations in employment and the unresponsiveness of wages.

## APPENDIX

*Computation of Distributions of Job Duration*

THE computations start from the published distributions of job tenure.<sup>33</sup>  
Let

$f(x, a)$  = fraction of the working population of age  $a$  who took their present jobs  $x$  years ago, as published.

The first step is to calculate

$G(x, a)$  = fraction of jobs taken by workers of age  $a$  that last no more than  $x$  years.

Now the fraction of the population of age  $a + x$  who took their present jobs  $x$  years ago is

$$\begin{aligned} f(x, x + a) &= \text{the probability of taking a job } x \text{ years ago at age } a \\ &\quad \text{times the probability that the job lasted until now} \\ &= f(0, a) [1 - G(x, a)]. \end{aligned}$$

Solve for  $G(x, a)$ :

$$G(x, a) = 1 - \frac{f(x, x + a)}{f(0, a)}.$$

Then let

$$\begin{aligned} g(x, a) &= \text{fraction of jobs taken by workers of age } a \text{ that last } x \text{ years} \\ &= G(x, a) - G(x - 1, a). \end{aligned}$$

Consider a worker with a typical work history. The task is to measure

$$h(x) = \text{fraction of a career spent in jobs lasting } x \text{ years.}$$

The typical worker will hold a number of jobs. If  $N$  is the total number of jobs,  $Ng(1)$  will last one year,  $Ng(2)$  will last two years, and so on. The length of the career will be

$$N \sum_{i=1}^n ig(i) = T,$$

33. Bureau of Labor Statistics, *Job Tenure of Workers, January 1973*, Special Labor Force Report 172 (Government Printing Office, 1975), table A, p. A-8.

**Table A-1. Computation of Distributions of Jobs by Duration, Workers Aged 30-34 at Beginning of Jobs<sup>a</sup>**

<i>Years in current job and corresponding age</i>	<i>Proportion of age group in jobs of this duration (1)</i>	<i>Interpolated density, f (2)</i>	<i>Cumulative distribution of jobs, G (3)</i>	<i>Distribution of jobs, g (4)</i>	<i>Proportion of career spent in jobs of this duration, h (5)</i>
½ or less	0.152	0.231	0.387	0.387	0.038
½-1	0.079	0.146	0.613	0.226	0.047
1-2	0.133	0.109	0.711	0.098	0.047
2-3	0.085	0.082	0.782	0.071	0.047
3-5	0.149	0.065	0.828	0.046	0.053
5-10	0.204	0.033	0.912	0.084	0.173
10-15	0.129	0.024	0.936	0.024	0.082
15-20	0.107	0.020	0.947	0.011	0.052
20-25	0.097	0.018	0.952	0.005	0.038
25-30	0.084	0.013	0.966	0.014	0.109
30-35	0.049	0.007	0.981	0.015	0.142
...	...	...	1.000	0.019	0.168

Source: Derived from Bureau of Labor Statistics, *Job Tenure of Workers*, table A, p. A-8.

a. Column 2 is computed from the published numbers in column 1 by linear interpolation. The density at zero is 0.377. Column 3 is computed from column 2 as  $G_i = 1 - f_i/f_0$ . Column 4 is computed from column 3 as  $g_i = G_i - G_{i-1}$ . This appears in table 3, column 3. Column 5 is computed by interpolating column 3 to form two values of the density corresponding to  $g_i$ , computing the mean duration (3.81 years), dividing the product of the duration and the density by the mean, and adding together to obtain  $h_i$ . This appears in table 3, column 4. Figures are rounded.

where  $T$  is the number of years in a working lifetime. Thus

$$N = \frac{T}{\sum_{i=1}^n ig(i)};$$

the number of jobs in a typical career is the length of the career divided by the mean length of jobs.

Now  $Ng(x)$  of the jobs last  $x$  years, so the fraction of the career in jobs of length  $x$  is

$$\begin{aligned} h(x) &= \frac{xNg(x)}{T} \\ &= \frac{xg(x)}{\sum_{i=1}^n ig(i)}. \end{aligned}$$

The details of the computations are given in table A-1.

## *Comments by Martin Neil Baily*

In both his own written work and his discussions of the work of others, Robert Hall has made outstanding contributions to the Brookings panel during the past ten years. In his paper for this meeting he has presented an incisive and fascinating analysis and critique of the theory of long-term wage and employment arrangements. I refer to this body of analysis here as contract theory, a theory that includes, but is not restricted to, the modeling of implicit wage contracts. The task is “to provide a solid economic rationale for the insensitivity of wages to current economic conditions and for the conspicuous deviations of employment from . . . trend.”

The basic starting point of contract theory is an attempt to reconcile the observed cyclical behavior of the labor market with the fundamental axiom that people act in their own economic interest. Such a focus may or may not reveal anything about market efficiency. That is usually a more complex issue. In addition, pure contract theory should be seen as an attempt to illuminate certain phenomena of the labor market, not to model every aspect of the complex reality observed. Of course, theory should be tested with the facts. But it is a question of how you do the scoring. Hall’s approach is to investigate how much can be explained by an efficient labor market model. That is a superb way to clarify the issues, but inevitably the fit between the theory and the facts will not be exact.

With the license given to discussants in this tenth anniversary edition to add their own view of the issues, I want to spend some time on the risk-sharing aspects of contract theory. This is important to Hall’s story—part of the glue that holds firms and workers together—but perhaps he gives too little attention to it. Human capital is the most important form of wealth for almost all blue collar workers and for most white collar workers also. There are remarkably few ways that workers can diversify the risk of their human capital. As Frank Knight noted many years ago, one of the roles of firms and entrepreneurs is to bear the risks of doing busi-

ness. And the ability of stockholders to diversify risks in the capital market enhances their ability to act as financial intermediaries for workers.

There are some difficulties that stand in the way of accepting the contract theory analysis of sticky wages as a form of insurance. First is the question of real versus nominal wage stickiness. It is alleged that contract theory predicts sticky real wages, whereas in fact nominal wages are sticky. For reasons I discuss below, contract theory need not stand or fall on this issue either way. But still it is worth checking the facts as a starting point. According to pre-World War II data presented by Albert Rees,<sup>1</sup> nominal compensation per hour in manufacturing in 1933 was only 84 percent of its 1930 level. In 1937 it was 144 percent of its 1933 level. This is hardly a pattern of total stickiness. By contrast, the real wage, computed using the consumer price index, was much more rigid. It rose or remained constant in every year of the Great Depression except 1932, when it fell only 1.3 percent.

In the postwar period the focus of the analysis seems to have slipped a derivative. Sticky wages are viewed in terms of sticky rates of change. I am not sure that is correct, at least for evaluating the contract theory approach, but I will go along with it. The rate of change of money wages has remained very sticky over some periods, for example, 1958–65, 1969–72, and 1976–79. These spells of inertia are of interest, but they do not settle the issue. Anyone who watches a roulette wheel over a prolonged period will observe runs where red comes up several times in a row. At various times, the rate of wage inflation has moved sharply. The index of compensation per hour in the nonfarm business sector increased 3.1 percent from 1953 to 1954. This rate of change almost doubled to 6 percent only two years later. A jump of similar magnitude occurred from one year to the next in the mid-1960s (3.5 percent in 1965 to 6.1 percent in 1966), and the rate of wage increase in 1975 was 50.0 percent above the 1969–72 figure. Clearly the rate of nominal wage increase can be quite volatile. Indeed, that is why econometricians have had so much trouble fitting Phillips curves.

The story for real wage increases, at least for the real wage computed from the consumer price index, is somewhat similar. There have been periods with rather stable trend increases of real wages and years of fluc-

1. Albert Rees, "Patterns of Wages, Prices and Productivity," in Charles A. Myers, ed., *Wages, Prices, Profits, and Productivity* (American Assembly, Columbia University, 1959), pp. 11–35.

tuations. Negative real wage changes have occurred very distinctively during periods when the components of the CPI such as housing, food, and energy have risen sharply. In the 1970s, for example, the sharpest declines occurred in 1974 and 1979, when the CPI jumped but the unemployment rate was 5.6 percent and 5.8 percent, respectively. These figures are below the average rate of unemployment for the decade. Over the entire period from 1954 to 1979, the variance of the rate of nominal wage increase (measured as percentage changes from year to year) was 4.3 percentage points, and the variance of the rate of real wage increase was 2.3 points. And even this rather low variance for the real wage is overstated. The CPI fluctuates when interest rates change, even when the true cost of living has altered very little. And the response of nominal wages to CPI changes is inevitably a lagged response.

Perhaps the real-nominal issue should be resolved by looking at the feedback terms in Phillips curves. I have found, as George Perry finds in his paper in this issue, that lagged wages lose significance in a Phillips curve with price feedbacks included. On balance, therefore, I do not regard the evidence as clearly favoring nominal over real wage stickiness. The key fact is that neither wage adjusts quickly in response to apparent excess labor supply. That is what is meant by stickiness.

I have discussed this empirical question at some length because it is raised as an objection to contract theory and because it is an issue of interest in itself. But, in fact, contract theory can readily accommodate nominal stickiness. Arthur Okun thought the wage-wage process was the basic one, and he was a major contributor to contract theory. The fundamental insurance a firm provides is that it will not use the *temporary* pressure that a recession creates to lower its workers' wages, either below their trend rate of growth or below their normal relation to the wages of other workers. This does not mean, however, that the real wage will be set in the long run at a level in excess of that justified by the underlying neoclassical forces. There is no guarantee that real wages will be fully protected against OPEC price increases or real estate speculation.

Some version of contract theory might predict that such a guarantee would be given, but it is not a version I recognize. Assume, first, that firms would have been able to anticipate the possibility of OPEC price increases, the anchovy disaster, the world grain shortage, the decline of the dollar, the decline of productivity growth, as well as all the other things that might have happened in the 1970s but did not. Assume, second, that

ex ante they would have been able to find assets that allowed them to diversify away all the risks. Only with these two absurd assumptions is there any case at all for the alleged prediction of guaranteed real wage growth in contract models. And even then the case is still weak. A fundamental property of implicit contracts is that one party will default when the *long-run* gains from compliance fall below the losses. This property suggests that wages cannot remain permanently out of line with their long-run market valuation.

Realistic contract theory recognizes the limits to the insurance coverage firms provide, just as there are limits to every other kind of insurance coverage. In normal times, real wages and relative wages are protected, even in a recession. When there are basic changes in the terms of trade of the industrial sector, these alter real wages. Ironically, some of the troubles in the 1970s have resulted because firms tried to provide too much insurance in the short run. The response of wages to the CPI had been built up before the shocks of the mid-1970s. The attempt by employers to provide some measure of real wage protection following OPEC price increases simply resulted in an escalation of overall inflation.

The second major question raised about contract theory concerns the variability of employment and hours—that is, the value of workers' time. Hall gives this issue a good deal of attention, but I would like to add my own perspective.

One of the strengths of contract theory is that it models job separations in a way that is entirely consistent with the observed fact that laid-off workers would prefer to continue working full time at the going wage; in other words, layoffs are involuntary. Workers *are* assumed to be aware that their employment situation is risky and that workers in unstable industries (like construction and durable goods manufacturing) receive higher hourly wage rates, other things being equal, than workers in stable industries (as indeed they do). But an enlisted man in the army who gets his head blown off would hardly be described as dying voluntarily, just because he knew ahead of time that army life was dangerous. Many people have reacted negatively to contract theory because they think that a model of efficient layoffs means unemployment is voluntary. I wish that issue could be laid to rest. When people react efficiently to adverse circumstances, it does not mean they like their situations.

A model of privately efficient behavior, as Hall expositis so clearly, indicates that layoffs take place when the value of the marginal product

evaluated at full employment is less than the value of the workers' time if laid off. Contract theory models can predict layoffs with or without postulating that leisure is valuable. There are two sides to the above inequality. If a firm in durable goods manufacturing were to maintain a full-employment level of output in a recession, it would have to cut its price to the point at which marginal revenue product was very low, perhaps zero, perhaps even *negative*. Product markets are not perfectly competitive, at least in the short run. Oligopoly, imperfect information (as in the Phelps-Winter formulation), or customer markets may all play a role in this imperfection.<sup>2</sup>

But for the record, I would assert that the value of workers' time when laid off, while it is less than the going wage, is also greater than zero.<sup>3</sup> Unemployment insurance is a principal reason for this in a modern economy. And some workers on short-term layoff supplement the value of this by enjoying their free time. If they do, more power to them. I would not conclude too much about any issue of economic policy from that. But for most workers who lose their "permanent" jobs for any length of time, unemployment checks become increasingly inadequate, and leisure hangs heavy. As a recent study of the unemployed concluded: "Virtually every jobless person I met supplemented unemployment checks with some sort of odd job or part-time work, a brand of illegality that seems as broadly accepted as petty income-tax evasion."<sup>4</sup> Other laid-off workers find temporary, stopgap, legal jobs while waiting for recall, and, of course, many workers do use their time to look for and find permanent, new jobs. There is no need to choose between contract theory and search theory; insights from both can be combined.

The choice of average weekly hours comes closest to the conventional analysis of substitution effects in labor supply. As Hall notes, a substantial fraction of the labor force consists of salaried workers who adjust the

2. See Edmund S. Phelps and Sidney G. Winter, Jr., "Optimal Price Policy under Atomistic Competition," in Phelps and others, *Microeconomic Foundations of Employment and Inflation Theory* (Norton, 1970), pp. 309–37; and Arthur M. Okun, *Prices and Quantities: A Macroeconomic Analysis* (Brookings Institution, forthcoming).

3. This statement refers to workers who expect to be rehired. Structural and long-term unemployment can create serious personal distress for the unemployed.

4. Harry Maurer, *Not Working: An Oral History of the Unemployed* (Holt, Rinehart and Winston, 1979), p. 6. The author is referring only to unemployed workers collecting unemployment insurance checks, and his sample was not scientifically selected.

intensity of their work effort to the pressure of work available as a matter of routine. For blue collar workers, variations in the amount of overtime follow cyclical demand changes. Reduced hours on the job are sometimes described as being of negligible value to production workers. This is not true. Turn the statement around and think about the disutility of overtime: workers like the extra income overtime brings, but not the overtime itself. The bitter International Harvester strike was largely over the issue of compulsory overtime. This suggests that the factory floor is not quite as much fun as some economists would have us believe. A reduction of weekly hours or a reduction of work pressure on the job is worth something to workers.

The third major question raised about contract theory concerns the incompleteness of the income insurance provided to workers who are laid off. Does the existence of this incompleteness undermine the model? I do not want to try and make a case that employer-employee relations reflect fully optimal insurance along every dimension.<sup>5</sup> When I read the insurance policies on my home, automobile, and health care, they do not look fully optimal. Kenneth Arrow has noted the rather odd limits and provisions of insurance coverage in medical care and other areas.<sup>6</sup> The basic wage stickiness theorem was shown in a model in which layoffs take place but full income insurance is not provided.

But do not exaggerate the extent of income insecurity facing workers with long-term labor contracts. The picture of the labor market Hall describes is one in which most employment occurs in jobs of long duration. With sticky wages, effective labor input (measured by output per employee) varies substantially, but workers spend a relatively small fraction of their time on layoff. This is a market, therefore, that provides considerable wage *and* job security to a large fraction of workers.

For those workers who are laid off, state unemployment compensation is a subsidized form of protection. A worker who spends ten out of fifty-two weeks on layoff and receives a 60 percent rate of replacement of lost earnings (allowing for taxes) will suffer only an 8 percent reduction of

5. In an earlier paper I argued that a two-tier labor market might be the relevant model, with upper-tier workers receiving close to an optimal contract but lower-tier workers receiving less. Martin Neil Baily, "Contract Theory and the Moderation of Inflation by Recession and by Controls," *BPEA*, 3:1976, pp. 585-633.

6. See, for example, Kenneth J. Arrow, "Foreword," in Howard Kunreuther and others, *Disaster Insurance Protection: Public Policy Lessons* (Wiley, 1978), pp. vii-ix.

annual income. Several million production workers in cyclically sensitive industries are eligible for supplemental unemployment benefits over and above the state systems.

Why is it that all laid-off workers do not receive full insurance? The basic reason is the tension between full insurance and incentives (or moral hazard) that occurs in any optimal insurance framework.<sup>7</sup> Firms cannot monitor their workers' behavior except when they are at work. It is neither privately efficient nor socially desirable for a firm facing a prolonged period of depressed demand to provide sufficient insurance that all its laid-off workers take a one-year vacation. Workers must be induced to find temporary or supplementary employment.

In addition, in practice there is not a hard-and-fast line between temporary layoffs and permanent layoffs. As the rational expectations theorists stress, a firm does not know precisely how much of a downturn is firm-specific, how much is industry-specific, and how much is economy-wide. In the Great Depression, for example, firms appear to have protected real wages to a remarkable degree. But most firms making layoffs feared that a large percentage of layoffs would be permanent. And many firms were concerned about their own survival. Efficient implicit contracts will not insure all of a firm's existing workers a constant income when a permanent reduction in the size of the work force occurs. Laid-off workers must be induced to find new, permanent jobs. Models that emphasize one kind of efficiency, as Hall's does, will inevitably neglect forces that yield another kind of efficiency—namely, efficient allocation of labor across firms.

Further, as Okun stressed, when firms and workers do not trust each other and when firms are themselves somewhat risk averse, the optimal partial insurance framework may be to use the "clean hands" policy of layoffs rather than to lower wages. The former has a self-enforcing property because firms lose the marginal products when they lay off workers.

The three main criticisms that are leveled at contract theory, therefore, seem to me to be answerable. Some of the answers may involve moving away from pure efficiency. But it is worth seeing how far one can go with

7. In the unemployment context see, for example, Martin Neil Baily, "Some Aspects of Optimal Unemployment Insurance," *Journal of Public Economics*, vol. 10 (December 1978), pp. 379–402; and Steven Shavell and Laurence Weiss, "The Optimal Payment of Unemployment Insurance Benefits Over Time," *Journal of Political Economy*, vol. 87 (December 1979), pp. 1347–62.

conventional economic tools; Hall shows it is possible to go surprisingly far.

I have one minor quibble. Hall is certainly right to emphasize the role of contract theory in understanding wages and labor input rather than using it as a theory of unemployment, but he overdoes it. The labor turnover data for manufacturing show the quantitative importance of layoffs and rehiring. During the initial quarters of a recession, like the one the economy is experiencing now, there are short-term plant closings and large numbers of layoffs of production workers. The trough of the recession is reached rather quickly, however, and firms rehire many of the laid-off workers. The layoff-rehire story is an important part of the response of the labor market to a sharp downturn. Moreover, the *indirect* impact of a sudden flood of laid-off workers on the employment prospects of new entrants, reentrants, and job changers is surely important.

Overall, this is an excellent paper. I like to think contract theory is an important area. I know Hall has contributed to the understanding of it. That understanding remains at a fairly primitive stage. But having a model of rational behavior in which the competition of workers for jobs does not determine the short-run path of the wage is, perhaps, a step toward explaining why the momentum of wage inflation is so hard to restrain by recession.

## *Comments by Lawrence H. Summers*

Many of Robert Hall's papers, including this one, have contributed greatly to our understanding of the functioning of the American labor market. A common thread in his work and that of many other economists during the past decade has been an effort to provide explanations for unemployment and economic fluctuations that have a sound microeconomic foundation. These explanations have almost universally tended to minimize the social cost of unemployment and the business cycle by suggesting that they represent an efficient response to changing economic conditions. Search theories, theories of intertemporal substitution, and at least some species of contract theory all imply that efforts to stabilize output fluctuations, even if successful, are likely to be counterproductive in welfare terms. I want to examine critically the specific equilibrium theory that Hall offers and comment on this general approach to understanding the labor market.

Hall's paper sympathetically but noncommittally outlines a theory of wage rigidity and employment fluctuations in an economy in which job attachment is near-permanent. It fills a major gap in knowledge by documenting the extent of job permanence in the economy. Hall finds that while the typical job is short, the typical worker is in the midst of a lengthy employment spell. According to his calculations, half of all employment occurs in jobs lasting over fifteen years. Because workers and firms both recognize the enduring nature of their relationship, the time path of compensation is determined by mutual convenience rather than market terms. The path of wages is dictated by considerations of income smoothing and insurance; it does not depend on the contemporaneous state of the labor market. Workers and firms negotiate efficient contracts—efficient in the sense that no alternative arrangement could make both parties better off. Efficiency requires that at all times the marginal product of labor and the worker's marginal rate of substitution between

income and leisure be equated. Because the former is more variable than the latter, the firm is permitted to determine the time path of employment subject to a set of contractual constraints.

Hall argues that this view of the economy can account for a wide variety of observed facts. He suggests that it provides an explanation for the striking wage rigidity observed during the past decade. He points to it as an explanation of the puzzling cyclical movements in productivity and of the procyclical movement of hours of work. Indeed, an important contribution of the paper is to focus attention on these two dimensions, which together account for about one-half of fluctuations in output. Finally, Hall notes that this type of theory rationalizes unemployment due to temporary layoff and may have something to offer about other sources of job loss.

In what sense can this sort of theory explain the type of wage rigidity observed during the last decade? The fact that it is nominal rather than real wages that have been rigid is a major embarrassment to most contract theories. Considerations of income smoothing and risk aversion by workers, which most contract theories stress, lead one to expect that contracts would fix real, not nominal, magnitudes. The amazing constancy of nominal wage inflation in the United States bears emphasis. During the 1972–79 period when the rate of inflation varied between 2.9 and 13.6 percent and the rate of unemployment between 4.8 and 8.9 percent, the rate of annual wage inflation varied within a 2 percentage point range. Real wages, on the other hand, fell by almost 5 percent in the past year and have consistently exhibited significant fluctuations. Note that the logic of contract theory implies that workers should be insured against all nominal shocks and most real shocks. Because workers are much more risk-averse than capitalists, and shareholders can diversify any firm-specific risk, wages would be almost fully indexed to a proper cost-of-living index if risk aversion were a dominant consideration.

There is a second important difference between the arrangements contract theories would lead one to expect and those observed in the real world. Considerations of consumption smoothing and risk aversion should lead to *income* rather than *wage* smoothing. Risk-neutral firms should insure risk-averse workers against income loss due to job loss or reduced hours. Indeed, this source of risk to a worker's real consumption is far greater than any uncertainty arising from wage flexibility. To some

extent, unemployment insurance meets this need. Yet replacement rates are typically much less than 100 percent, and this insurance does not cover the underemployed at all. Although a small number of union contracts provide for supplemental unemployment benefits, the majority of employment contracts make no provision for such income smoothing. It might be argued that workers have other sources of income such as stopgap jobs, so there is little need for compensation from the firm. Note that efficiency implies that a worker's consumption should not be affected by being laid off. This would imply that workers should envy those who are laid off, and have the same consumption and more leisure. I doubt that this implication is met in practice.

Hall's theory suggests that explaining movements in wages is of secondary interest because the contemporaneous wage has no important effects. Pursuing this line a little further leads to another difficulty. If the observed wage bears little relationship to the current market condition, there is no reason why it should be as tightly linked to output prices as it appears to be in the data. If goods are sold competitively, then in Hall's equilibrium world, price should always equal the marginal cost as measured by the value of a worker's time. According to this view, strongly countercyclical movements should be expected in observed real wages, since prices should fluctuate freely and nominal wages do not. In fact, real wages are not cyclical. Indeed, it is precisely this paradox that has spurred much of the research on wage rigidity. Semi-permanent contracts between buyer and seller might help to explain the sluggishness of prices, but they do not appear to be features of more than a few output markets.

The best that can be said for the "efficient fluctuations" model is that it is not obviously refuted by the rigidity of wages and prices. It certainly does not predict the types of rigidity observed. This may not be of great importance because Hall's theory is really one of wage irrelevance. What is crucial is the suggestion that employment fluctuations are efficient because the value of the marginal product of labor is equated with the workers' marginal rate of substitution. Here Hall recognizes that to tell a convincing story he must have fluctuations in value of the marginal product of labor. For a single-firm industry this is easily done by postulating a decline in the relative price of its output. Hall forcefully points out that this sort of thing cannot be simply extended to explain fluctuations throughout the economy. This inability to explain economy-

wide movements in employment seems to be a fatal flaw in the efficient fluctuations model. Its irrelevance to aggregate economic performance calls into question even its relevance to micro fluctuations.

Following Hall in focusing on the supply side, I next examine the consequences of a decline in the marginal product of labor, without worrying about its source. He argues that firms will slide workers down their supply curves, achieving efficiency along the way. Some workers will be laid off; more will face reduced hours and supply far less effort. The argument that it is efficient rests on the premise that workers place a high value on brief periods of work reduction. The implied high short-run elasticity of labor supply is a lynchpin of almost all equilibrium theories of the business cycle. Yet it lacks direct empirical support. Most empirical studies assume what is at issue: that at each point in time no workers are rationed, and so all are on their supply curve. The only microeconomic study that has attempted to relax this assumption, done by Orley Ashenfelter, found evidence that many people were in fact unable to "sell" as much labor as they wanted. This is true at all stages in the business cycle. There is other available evidence which suggests that intertemporal substitution effects are likely to be small. The labor supply response to temporary changes in marginal tax rates or to announcements of future changes in the permanent tax rate does not appear to have been large. If intertemporal substitution effects were genuinely important, the decline in the trend growth in real wages during the past decade should have had profound implications for patterns of lifetime labor supply. These patterns do not appear to have occurred. Finally, there is the direct testimony of the affected parties. The unemployed report that they want to find work at prevailing wages. Survey data reveal that a large number of workers say they would prefer more hours or weeks of work at prevailing wages, if they could get them. These considerations lead me to doubt that the efficient response to even quite large declines in demand would give rise to a significant decline in labor input. At a minimum it would seem that the burden of proof is on proponents of this hypothesis.

Hall recognizes that even if demand shocks are assumed and a high intertemporal elasticity of supply is postulated, his theory cannot account for more than a small fraction of cyclical unemployment. He notes that only about 12 percent of the unemployed are classified as temporary layoffs. Even this figure greatly exaggerates the number of unemployed

workers who fit into his framework. Those who are laid off search for work almost as intensively as the remainder of the unemployed. This suggests that they may not regard their job attachment as permanent. Nor should they. Data from the Current Population Survey demonstrate that fully half of those reported as being on temporary layoff do not even return to their original occupation and industry. Finally, it is important to note that a significant part of changes in the number of persons working is accounted for by fluctuations in the size of the labor force, which have nothing to do with workers with permanent job attachments.

Hall suggests that the real relevance of contract theories is in explaining movements in labor input per worker; that is, changes in hours and effort. To some extent those movements must have a voluntary component. I am skeptical, however, about its importance. A significant fraction of movements in hours per worker can be accounted for by changes in the number of persons reporting themselves as involuntarily working part-time. An additional component is due to changes in the amount of overtime work which, given the sharp kink in the budget constraint at forty hours, are likely to be partially involuntary. It would surely be inappropriate to attribute most of the cyclical increments in productivity to intertemporal substitution of work effort. As Arthur Okun showed, industry shifts account for a significant part of the aggregate productivity effect: a large portion of the remainder must be attributed to increasing utilization of overhead labor. Robert J. Gordon demonstrated that some of the procyclical productivity pattern is attributable to lags in employment adjustments following shocks. Because effort cannot be measured directly, it is difficult to evaluate the intertemporal substitution claim in this area.

One other difficulty deserves mention. It does not seem to me that a theory like the efficient contract theory can be used to explain some but not all aspects of labor market behavior. If any inefficiencies are present, the theory of the second best teaches us that it is inefficient to meet the normal optimality conditions or other margins. Hence in the presence of involuntary unemployment it would no longer be efficient to meet Hall's conditions for hours and effort. They would not be satisfied by efficient contracts that recognize the possibility of disequilibrium elements in the labor market.

Hall's concluding comments could be taken to suggest that we have a better understanding of the supply side of the labor market than the

demand side. This statement is correct only if models are considered in which the labor market is in permanent equilibrium, where there is no explanation of all the demand fluctuations, but there are conjectures about the supply side. I would describe our knowledge quite differently. A few things are known for sure: nominal wages move very sluggishly and employment fluctuates substantially. At certain times the number of people who say they want jobs at prevailing wages but cannot find them rises sharply. At these same times, many firms report that they cannot sell all their output, so they reduce their inventories and shut down their factories. As Hall acknowledges, these events tend to follow money contractions. Equilibrium theorists dismiss the contemporaneous wage as irrelevant, attribute apparent unemployment to misperceptions about the actual wage level, and ignore the intermittent idling of the capital stock. They may be right. But it seems to me more promising to try to explain the facts rather than to explain them away. If the sluggishness of nominal wages is postulated, the whole thing becomes quite explicable within the "mainline" model of inflation. Only in the long run do demand shocks fall primarily on prices. Given the rigidity of nominal wages, demand shocks will have large effects on output and employment. Shortfalls in demand give rise to large social costs in terms of forgone output. Any policies that can reduce these shortfalls are desirable. This view of the world accommodates all the three embarrassments to the efficient contract theory that Hall isolates. With rigid wages, demand fluctuations and large monetary effects are to be expected. Substantial involuntary unemployment is to be expected in markets where prices adjust slowly. So are fluctuations in hours and productivity. Both can be seen as the natural response of firms that find themselves temporarily constrained in the output market and that face costs of labor turnover. In a world of nominal rigidities and sociological constraints, high natural rates of unemployment also become more explicable, as firms cannot readily bargain with individual workers.

There is, of course, one huge difficulty in all of this. Given the enormous social costs engendered, why should nominal wages move so sluggishly? As Robert Solow has recently argued, there is unlikely to be a single explanation, and the multiplicity is apt to involve as much sociology as economics. It seems to me that understanding the fact of nominal wage rigidity is the major unsolved problem in understanding the

labor market. This seems to me more fruitful than trying to demonstrate its irrelevance.

Hall's paper does offer one lead here. The permanence of job attachment that he documents implies that most workers are not likely to face the threat of prolonged unemployment, even in recessions. This is especially true for workers with significant seniority. Almost all employment adjustment includes persons who have held jobs only briefly; over three-quarters of cyclical fluctuations in employment involve secondary workers. These facts may explain why unemployment has so little impact on wages. The labor market is really many markets. In the markets where most wages are paid, there is essentially no unemployment. Even a large fall in the wages of the potentially unemployed would have a small effect on overall wages. It is this dual nature of the labor market that may account for the flat short-run Phillips curve.

In sum, Hall has presented a fascinating paper. It is the most successful effort to date to mesh equilibrium theories with the realities of the labor market. At the same time, it emphasizes the severe empirical difficulties that plague equilibrium theories. If in the end one is left unconvinced by the equilibrium viewpoint, the failure is in the theory itself, not in Hall's ingenuity.

## *General Discussion*

Several panelists elaborated on the issue of wage inertia. James Duesenberry observed that most adult male workers were committed to working full time so that the labor-leisure trade-off was not particularly relevant in explaining their behavior. He argued that, rather than trying to model the employment-wage decision in terms of efficiency and marginal conditions, it should be modeled as a problem in managing a work force with procedures aimed at recruiting and retaining the needed quality of workers. Under most conditions, firms accomplish this by maintaining relative wages in the face of fluctuations in their demand for labor. The cost of living could enter wage-setting in a minor way if periods of disappointing real wage growth provoke workers to reconsider the decision to stay with an employer as opposed to looking for a better job. Franco Modigliani reasoned that the intended compensation path might be defined in terms of real wages but that less than immediate adjustment to unanticipated inflation would result in fluctuations in actual real wages. William Nordhaus added that there are substantial costs and problems associated with writing contracts that fix the real wage. Writing nominal contracts is a suitable way of avoiding these costs when the inflation rate is fairly steady, as it had been in the United States before the mid-1960s. In time, however, more variable inflation led toward greater indexation of U.S. wages. James Tobin reasoned that the theory of long-term employment arrangements did not imply contracting would fix either the nominal or the real wage; instead, it led to the vaguer concept that wages would be governed by expectations of fairness and meeting wage norms. What satisfies these expectations or determines these norms has not been established in economic models and might vary with time.

Some panelists commented on the issue of macro efficiency, which Robert Hall's paper separated from the issue of micro efficiency in employment arrangements. Robert Gordon noted that there is no puzzle in

explaining fluctuations in aggregate employment if nominal wage stickiness is accepted and nominal demand is given, although this leads to a disequilibrium description of total output and employment. Nordhaus noted that such a disequilibrium view of macro fluctuations explained the actual facts of the typical cycle considerably better than the equilibrium business cycle view associated with Robert Lucas and Thomas Sargent.

Tobin pointed out that Hall's view of long-term employment arrangements provided a rationale for the existence of a collective bargaining process. First, the costs of turnover to the firm rise nonlinearly, so that collective action is useful to workers. Second, there will be a gap between the marginal product of labor and the reservation wage that reflects the marginal value of the worker's time, and this gap provides a zone within which a bilateral bargaining process can occur. Tobin also noted that normal attrition through voluntary quits and retirements provides firms with considerable flexibility in determining employment within the context of the model without resorting to layoffs.

Peter Kenen stressed that, even if one accepted Hall's contractual model as consistent with equilibrium behavior at the level of the firm, the model could not describe an equilibrium for the entire work force because many of its members did not have access to contractual jobs. As evidence against the equilibrium interpretation of layoffs, Thomas Juster cited survey results, which showed that workers did very little with their time when they were laid off. Juster was skeptical about the accuracy of Hall's data on average job length that was based on people's memories. He suggested pursuing the question with a more reliable data base.

Robin Marris noted that firms in other economies responded differently to demand variations than firms in the United States. In Europe and Japan there typically has been more work-sharing and a greater flexibility in compensation achieved by changing overtime and bonuses. As a consequence, production has declined without a corresponding drop in employment. As white collar employment becomes more prominent in the United States, total employment fluctuations may move toward the pattern observed elsewhere.