

MARTIN NEIL BAILY

Yale University

*Contract Theory
and the Moderation
of Inflation
by Recession and
by Controls*

THE EXTRAORDINARY persistence of price inflation in the face of mild and then severe recession over the last several years has suggested a momentum in wages and prices. Once inflation has become the pattern, it is very hard to eliminate. One view of this momentum is that the historical experience of rising wages and prices creates the expectation of further inflation; and, because firms and workers expect prices and wages to rise, they do rise. Provided monetary-fiscal policy is accommodative, these expectations are fully realized and the inflation continues. Another, and not entirely distinct, view holds that firms set wages and prices relative to other wages and prices. The inflation continues not so much because of expectations, but because firms and workers constantly perceive themselves as just catching up with past inflation.

There is broad agreement that refusing to provide fiscal-monetary accommodation will eventually end an inflationary process. Whether monetary or fiscal restraint is the tool is not an issue here. In either case, the

Note: I would like to thank Ray C. Fair and M. A. Baily, participants in the Brookings panel, and numerous other colleagues for helpful comments and criticisms. D. S. Coppock provided excellent and invaluable research assistance.

result is a period of recession, of underutilized labor and capital. In particular, the unemployment rate will exceed for a period the rate sometimes called "natural," but more appropriately called the nonaccelerating-inflation rate of unemployment (NAIRU) since this rate is neither "natural" nor optimal. The momentum of inflation is a particularly severe problem for the macroeconomic policymaker because halting it by restraining aggregate demand becomes so very costly. Why a severe or prolonged recession seems necessary to halt inflation has long been a central issue of macroeconomic theory and poses the main dilemma of current macroeconomic policy.

This paper will consider certain aspects of this question, but before summarizing the main ideas, it is worth clarifying the viewpoint that will be taken on two controversial issues. The first concerns the sharp disagreement about the chain of causality by which monetary or fiscal restraint eventually slows inflation. One hypothesis asserts that policy changes have a direct impact on real aggregate demand, which, in turn, has a direct impact on real aggregate output and on employment. Given an initial fall in demand, unemployment will rise above the NAIRU and then the rates of wage and price inflation will decline. An alternative hypothesis is that changes in policy have a direct impact on prices, affecting output, either because of short-run money illusion or, equivalently, because the actual and the expected price levels diverge. It is not clear which, if either, of these views should be labeled "Keynesian" or "classical." Keynes assumed a type of money illusion¹ and so do many modern textbooks. But errors in the perception of the price level are a feature of the new classical macroeconomics, allowing such models to display "'Keynesian' short-run properties."² Many of the arguments made in this paper apply in either framework, but at various points it will be assumed that the causal chain runs from policy to real output to wages and prices.³

The second controversial issue concerns the way in which expectations

1. John Maynard Keynes, *The General Theory of Employment, Interest and Money* (London: Macmillan, 1936), especially book 5.

2. Thomas J. Sargent, "A Classical Macroeconometric Model for the United States," *Journal of Political Economy*, vol. 84 (April 1976), p. 235.

3. A detailed critique is not appropriate here, but two of the arguments that tip the scales against the alternative view are (1) that the real wage does not show the appropriate countercyclical movement; and (2) that the profit-maximizing output of a firm depends upon its own input and output prices and not upon the general price level.

are formed. The new classical models assume rational expectations, ensuring that any Keynesian properties they have indeed apply only to the short run. The assumption also means that only unexpected policy changes or exogenous shocks will cause actual prices to deviate from expected ones and hence cause an unemployment rate different from the NAIRU. It is possible, however, to accept the assumption of rational expectations and reject other assumptions of the new classical models. In fact, many of the contract-theory models built by myself and by others assume rational expectations. However, the assumption of rational expectations, while convenient analytically, is probably at best only an approximation to the rather casual way in which people actually form expectations, heavily influenced by their own direct experience. As a result, different groups may hold very different expectations. This paper at times makes assumptions inconsistent with the usual statements of rational expectations. However, the implications of more “rationality” will be considered, as well as whether the label is always being correctly applied.⁴

Two related ideas will be presented. First, it will be argued that a world characterized by wage and price setters, and in particular implicit or explicit contractual wage and price relationships, will display the observed property of short-run changes in output and employment and slow wage and price adjustment. Second, in such a world, it is possible, at least in principle, to design a “request-control” system of wage and price controls that could slow inflation without the loss of real output resulting from demand-restraint policies and the restriction on relative-price adjustments characteristic of earlier controls. The analysis of the response of the economy to recession emphasizes the labor market. Equal time will not be devoted to product markets. This focus reflects both my comparative advantage and my belief that the response of the labor market is very important in the overall picture. The product side is considered, however, along with labor markets, when controls are analyzed and when the possibility is recognized that fluctuations in prices of raw materials may cause price inflation to diverge from its usual relationship to wage inflation.

The paper starts with a simulation model that shows how the ability of a given recessionary path of output to slow the rate of inflation is diminished, *ceteris paribus*, when wages are adjusted less frequently because wage contracts are in force. This result is a partial one. It does show

4. Certainly, many financial markets satisfy efficiency or random-walk properties. Price-level expectations are not subject to the same discipline of arbitrage.

that wage contracts can be an important determinant of the economy's response to recession. But it does not explain (1) why such contracts exist, or (2) whether other things are significantly different because contracts exist. The following two sections evaluate some efforts to develop a theory of implicit and explicit contracts. Some extensions of the theory are then presented, along with the argument that the new theory provides at least the outlines of an explanation for the existence and nature of wage contracts, one that is firmly based upon established economic principles. The next two sections consider the alternative method of slowing inflation by wage and price controls. In the highly stylized economy considered first, the request-control system would be perfectly effective in stemming inflation without imposing distortions at the microeconomic level. Some issues that would arise if the system were applied in a more realistic setting are then considered. The discussion of the control system can be viewed in two ways: as an outline of a potentially practical method of applying controls or alternatively as a useful exercise in theory against which to measure actual control systems.

Simulating the Effect of Wage Contracts

In preparing his presidential address to the American Economic Association,⁵ James Tobin developed a simulation model that emphasized the behavior of the labor market and disaggregated it into ten sectors.⁶ In this model, real aggregate demand is determined exogenously although the implied IS-LM determination would not be hard to specify. A given level of aggregate demand is divided among the sectors according to a stochastic process, lending the model the desirable feature of a stochastic equilibrium in which some sectors have excess demand for labor and some have excess supply. Past wages or prices feed back into current wage increases in each sector. Those sectors with excess demand for labor will tend to accelerate the rate of wage inflation and those with excess supply will tend to decelerate it. The stochastic equilibrium occurs when these pressures just offset each other except for random movements. Unem-

5. James Tobin, "Inflation and Unemployment," *American Economic Review*, vol. 62 (March 1972), pp. 1-18.

6. I have raised the number to twelve, because that is divisible conveniently by two, three, four, and six.

ployment exists in equilibrium—a NAIRU, in fact—because workers in excess-supply sectors cannot immediately fill vacancies in excess-demand sectors. Workers do move between sectors, however, in response to wage differences and vacancies.

This model looked like a suitable framework within which to explore the impact of changes in the frequency of adjustment of wages. In the Tobin specification, wages are changed in every sector in every period (here, a quarter). The effect of contracts was then simulated by introducing annual wage changes and, in a multiyear contract, by making the amount of the increase (or decrease) the same in each year of the contract.⁷ This procedure will not make wages stickier in any trivial way because the specified annual increase is four times the quarterly increase implied by the Tobin parameters. Contracts simply slow down the frequency of wage adjustments. However, since the recession is specified as a given path of real aggregate output, the reduced frequency of adjustment will actually curtail the reduction of the inflation resulting from the given recessionary path. This is the main result. Furthermore, using this model lends some objectivity to the results that would be lacking in those of a simulation model that I made to order for this specific purpose.

FORMAL STRUCTURE

Real demand in each sector is set according to

$$(1) \quad \begin{aligned} D_{i0} &= a_0 && \text{for all } i \\ D_{it} &= \frac{a_t}{a_0} [D_{it-1} (1 + \rho) + \epsilon_{it} (1 + \rho)^t] && \text{for all } i \\ \sum_i \epsilon_{it} &= 0 && \text{for all } t, \end{aligned}$$

where D_{it} is the demand for the output of the i th sector in period t , ρ is the exogenously set quarterly rate of increase of productivity (3 percent annually), and ϵ_{it} is a random variable chosen by a random-number generator. Real aggregate demand is set by choosing a_0 , the initial level in each of the sectors, and then the path of a_t over time; $a_t = a_0$ simply holds the quarterly rate of growth in aggregate demand at ρ . This is the trend

7. Susan J. Lepper gave important assistance in converting the Tobin model to allow contracts. She has studied the effect of indexing on wage contracts using this model; see her "Wage Indexing: Boon or Boom?" (preliminary paper prepared for the Econometric Society meetings, 1974; processed).

growth rate of potential output for the economy since labor-force growth and capital accumulation are suppressed. If a_0 is chosen so that the economy starts in stochastic equilibrium, then setting $a_t < a_0$ for a period will induce a recession.

Producers are assumed to set a desired level of output by extrapolating last period's demand by the trend rate of growth of demand and adding a fractional adjustment of actual inventories to desired inventories:⁸

$$(2) \quad Q_{it}^* = [D_{it-1} + b(S^* - S_{it-1})](1 + \rho),$$

where Q^* is desired output, S^* desired inventory, and S actual inventory. Actual output (Q) is the smaller of desired output and capacity output (C):

$$(3) \quad Q_{it} = \min [Q_{it}^*, C_{it}].$$

And capacity output is determined by the available labor supply (L) adjusted for productivity growth:

$$(4) \quad C_{it} = L_{it}(1 + \rho)^t.$$

The above relations are then used to determine the excess demand or excess supply for labor in a particular sector:

$$(5) \quad X_{it} = Q_{it}^*(1 + \rho)^{-t} - L_{it},$$

where X_{it} is the excess demand for labor in sector i and is positive whenever desired output exceeds actual output and negative whenever desired output is less than capacity output.

The total labor force is constant but moves between sectors in response to differences in wages (W) and job opportunities:

$$(6) \quad Z_{ijt} = e \left(\frac{W_{it} - W_{jt}}{W_{it} + W_{jt}} \right) + f(X_{it} - X_{jt}),$$

where Z_{ijt} is the number of workers moving from sector i to sector j in period t . This means labor supply is given by

$$(7) \quad \begin{aligned} L_{i0} &= a_0 \\ L_{it} &= L_{it-1} + \sum_{j \neq i} Z_{ijt-1}. \end{aligned}$$

In a sector with no wage contracts the adjustment in each quarter (\dot{W}_i) depends upon the magnitude of excess supply or demand in the current quarter and its average over the four preceding quarters plus a term repre-

8. Desired inventories are constant, equal to 100 units of real output.

senting the average rate of price inflation (\dot{P}) in the four preceding quarters:

$$(8) \quad \dot{W}_{it} = cX'_{it} + d\bar{X}'_{it} + 0.25 \sum_1^4 \dot{P}_{t-k} + \rho,$$

where

$$(9) \quad X'_{it} = \begin{cases} \frac{X_{it}}{\sum_{k=1}^5 L_{it-k}}, & X_{it} < 0 \\ \frac{2X_{it}}{\sum_{k=1}^5 L_{it-k}}, & X_{it} > 0 \end{cases}$$

so that the effect of excess demand is twice that of excess supply. The variable \bar{X}'_{it} is defined similarly, replacing X_{it} above with the average of X_{it} over the four quarters $t - 1$ to $t - 4$. The wage level in each sector is then given by

$$(10) \quad W_{it} = W_{it-1} (1 + \dot{W}_{it}),$$

and then the price level in each sector is set by a markup relation:

$$(11) \quad \dot{P}_{it} = \dot{W}_{it} - \rho.$$

The average rate of price or wage inflation is computed as a weighted average of the sectoral rates using a chain index with the output of each sector as weights.

Equations 8 and 11 imply a wage-price spiral but substituting 11 into 8 could just as easily turn the model into a wage-wage spiral. The distinction between the two cases is, of course, crucial in the real economy, which is subject to fluctuations in commodity prices that break down equation 11, but I do not deal with this problem. To summarize, the main properties of the adjustment equation are (1) current excess supply or demand has a bigger influence than past conditions (actually $c = d$ in equation 8, but \bar{X} is an *average* of past quarters); (2) wages respond twice as rapidly to excess demand as to excess supply; (3) there is a full feedback of prices (or wages in the wage-wage case); and (4) with no excess supply or demand and no price inflation, wages grow at the same rate as productivity.

SIMULATION RESULTS

Wage contracts of one, two, three, and four years were considered in the model. In all four types of contract, wages are changed annually, with the same annual increase applying to each year of multiyear contracts. The amount of the annual increase in contract sectors was based on the parameters of equation 8; hence, the quarterly rate of increase from 8 was multiplied by four.⁹ Three combinations of contract lengths were analyzed as distinct cases, and a no-contract situation formed the fourth case. The three combinations were (1) eight sectors of two-year and four of one-year contracts; (2) all twelve with three-year contracts; and (3) eight with four-year and four with two-year contracts. The contract renewal dates were staggered evenly in all cases.

To examine the response of the model to recession, simulations were run comparing the rate of inflation obtained with steady growth of real aggregate demand to the rate resulting from a fairly long and severe recession. The "no recession" runs assumed that real aggregate demand grew at the trend growth rate of potential output throughout. The initial level of demand was set so that the unemployment rate was about 3 percent except for random changes; and it remained close to that initial rate in the no-recession case, as shown in column 1 of table 1. The model with quarterly wage adjustments—that is, no contracts—was first simulated with an initial price inflation rate of 12 percent and no recession. The resulting annualized rates of price inflation, shown in column 3 of table 1, drop slightly at first; it then follows a path with no particular tendency to accelerate or decelerate, responding to random shocks and displaying positive serial correlation because of the wage-price feedback.

The simulation runs with recession had a path of real demand that differed from the no-recession path only during quarters 24 through 58, when they specified a rather severe decline in real demand, followed by a recovery. The resulting path of unemployment is given in column 2 of table 1. The effect of this recession on the rate of price inflation in the no-contract version of the model is shown in column 4 of table 1. The differences between the recession and no-recession inflation rates, given in column 5 of table 1, reveal the considerable impact of the recession. The reduction in the inflation rate by the thirteenth year (just over halfway

9. This technique is not precise because it does not allow for compound-interest effects.

through the recession) is over 12 points. As the path of real demand returns to the no-recession path, the model settles down with inflation reduced by 13.8 points.

For each of the contract cases, the same simulation runs were made to compare the path of the inflation rate without and with recession. The arithmetic differences for each of the cases are shown in columns 6, 7, and 8 of table 1. The amount by which the given recession reduces the rate of inflation is much smaller for all of the contract cases than for the no-contract case, and is progressively smaller as the contracts lengthen.¹⁰

INTERPRETING THE RESULTS

A simple technical or mechanical reason within the model explains why an increase in the contract length dilutes the effect of a given recession on inflation. It is because the longer contracts are relative to the duration of the recession, the less frequently is the rate of wage *increase* from equation 8 subjected to the downward pressure of excess supply. A more behavioral explanation can be offered along the following lines. First substitute equation 11 into 8 to give the wage-wage version of the model:

$$(8') \quad \dot{W}_{it} = cX'_{it} + d\bar{X}'_{it} + 0.25 \sum_{k=1}^4 \dot{W}_{t-k},$$

where \dot{W} is the average rate of wage inflation across all sectors. According to this equation, workers and firms determine the appropriate nominal wage increases for their sector over the contract life from the recent experience of excess supply or demand and the recent rate of economy-wide wage inflation. Consider a sector that comes up for renewal when the average rate of wage inflation has been 10 percent. Suppose that the remaining terms of equation 8' (involving the X s) add up to -1 percent. If this sector has a one-year contract, it will set a single wage increase of 9 percent. Compared with wage increases granted in recent contracts, the workers settle for a 1 percent decline in their relative wage over the one year of the contract. If the sector has a three-year contract, it sets a wage increase of 9 percent in each of the three years. With the same expectations, the firm and its workers settle for a 1 percent decline in the relative

10. The average length of contract in the one- and two-year case is one and two-thirds years. The average length in the two- and four-year case is three and one-third years.

Table 1. Simulation of the Impact of a Recession on the Inflation Rate with and without Wage Contracts

Year	Unemployment rate (percent)		Inflation rate in no-contract case (percent)		Difference in inflation rate as a result of the recession (percentage points)			
	Without recession	With recession	Without recession	With recession	No contracts	One- and two-year contracts	Three-year contracts	Two- and four-year contracts
	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
1	3.4	3.4	11.7	11.7	0.0	0.0	0.0	0.0
2	3.2	3.2	11.4	11.4	0.0	0.0	0.0	0.0
3	3.2	3.2	11.2	11.2	0.0	0.0	0.0	0.0
4	3.3	3.3	11.0	11.0	0.0	0.0	0.0	0.0
5	2.9	2.9	10.9	10.9	0.0	0.0	0.0	0.0
6	3.2	4.0	10.8	10.7	0.1	0.1	0.0	0.1
7	2.9	7.4	10.7	9.6	1.1	0.7	0.3	0.2
8	3.0	10.0	10.7	7.8	2.9	2.0	0.9	0.6
9	3.3	9.8	10.7	5.7	4.9	3.2	1.5	1.2
10	3.2	9.7	10.7	3.7	7.0	4.5	2.1	2.1
11	3.0	9.7	10.8	1.7	9.1	5.9	2.9	2.7
12	2.8	9.7	10.8	-0.3	11.1	7.0	3.4	3.3
13	3.2	7.8	10.9	-2.0	12.9	8.1	4.3	3.6
14	3.1	4.6	10.9	-2.9	13.8	8.5	4.5	3.9
15	3.0	3.0	10.9	-3.0	13.8	8.9	4.7	4.1
16	3.2	3.2	10.8	-3.0	13.8	9.0	4.7	4.2
17	3.5	3.5	11.0	-2.8	13.8	8.9	4.7	4.3
20	3.0	3.0	11.1	-2.7	13.8	9.2	4.8	3.8
25	2.9	2.9	11.4	-2.4	13.8	9.1	4.6	4.0

Source: Based on simulation model discussed in the text. Figures are rounded.

wage over *each year* of the contract, about a 3 percent decline overall. Thus, relative to the expectations held at the time the contracts are made, the wage-setting decision process is no stickier in the three-year than in the one-year case. But the result is different when the actual rate of inflation is declining. Then, a high rate of expected inflation is incorporated into a three-year contract, whereas three one-year contracts will reflect annual revisions in expectations.

In addition to the rate of inflation, the conditions of excess supply or demand are also extrapolated for the length of the contract. The impact on the simulation runs is not great, however, because the decline into the recession is reversed during the recovery. Some contracts do not come up for renewal for several quarters after the recession hits. Thus, the

initial response of the economy reflects the wage increases set prior to the development of general excess supply. On the other hand, as the economy moves out of the recession, there will occur wage increases that were set during its course. These forces may not offset each other exactly, but they certainly do so at least partially.

The robustness of the finding of a smaller reduction in inflation with contracts may be subjected to criticism on two counts. First, extrapolating the recent rate of inflation is a naive way of forecasting. One might argue instead that once a decline in the rate of inflation is experienced, firms and workers would likely anticipate further declines and hence would set declining wage increases over a three-year contract; such a process could reduce or even eliminate the differences shown in table 1. In defense of simple extrapolation, the behavior of equation 8 may be realistic, even if naive, and furthermore may reflect a decision rule that is not so naive in practice. The actual behavior of wage inflation (measured annually by the rate of change of the index of adjusted hourly earnings in the private non-farm economy) gives firms and workers no reason to expect that a decline in wage inflation will be followed by a further decline. If anything, firms and workers may have learned that an initial decline tends to be followed by an increase in the following year.¹¹ Of course, extrapolation does not use structural information, as full rationality requires. Yet economists have not done all that brilliantly in forecasting wage inflation using structural information. When the policy decision was made in 1974 to fight inflation by means of recession, firms and workers who did not immediately build into their contracts the expectation of declining inflation were not necessarily irrational.

The second source of doubt about the results of the simulation concerns indexing. If the wage contracts were indexed, the differing contract lengths in the model would certainly have less impact.¹² However, few contracts

11. *Employment and Training Report of the President, 1976*, table C-10, column 4. The data are annual and are adjusted for overtime and for interindustry employment shifts. The period was 1950–75. If the rate of wage inflation ($\Delta W/W$) is regressed on its own lagged value and the change in the inflation rate, the resulting equation is $(\Delta W/W)_t = 1.02 (\Delta W/W)_{t-1} - 0.20 [(\Delta W/W)_{t-1} - (\Delta W/W)_{t-2}]$, with standard errors of 0.048 and 0.18, respectively, and an R^2 of 0.52. The difference term is not significant and the point estimate of the coefficient is negative. Regressing the rate of inflation on its lagged value alone yields $(\Delta W/W)_t = 1.01 (\Delta W/W)_{t-1}$, with standard error of 0.047 and R^2 of 0.50. One cannot say that experience has overturned a simple extrapolation rule.

12. See Lepper, "Wage Indexing," for further analysis.

in the United States are indexed; and those that are often are asymmetrical, responding little to decelerating inflation.¹³ The reasons for the infrequency of indexing are complex. The cost of living is only one of the variables that firms and workers consider; others include wages elsewhere and the firm's expected product price. No way of indexing wages offers a clear improvement over basing nominal increases on expectations over all the relevant variables.

While I am prepared to argue in defense of the general formulation of behavior in the model, I would stress that the simulation results should be regarded as very partial and strictly *ceteris paribus*. It is possible that varying contract lengths will induce other parametric changes that would offset the impact shown by the model. At most, the model provides a *clue* to the importance of wage contracts in macro dynamics.

OTHER FEATURES OF THE MODEL

The model illuminates some other important issues. First, longer contracts show greater resistance to the acceleration of inflation as well as to its deceleration. This is a plausible result, which may offer a clue to why the mid-1960s saw rapid growth, declining unemployment, and only slight acceleration of inflation. Second, for the reason just given and for reasons discussed later, contracts are not necessarily a "bad thing" in welfare terms. A contract comes into existence because the parties believe it will raise welfare and reduce costs.

Third, although the differences in contract lengths mainly affect the response to recession in the model, there are also small changes induced in the steady-state or no-recession simulations. According to the characteristics of labor mobility in the model, specified by equation 6, workers move in response to differences in excess supply or demand and in wages in different sectors. The efficiency with which labor is allocated is important. If some omniscient central planner were to move workers immediately from one sector to another in response to the random shocks, he could eliminate unemployment, avoid any excess supply of or demand for labor, and prevent acceleration of wage or price inflation. The actual allocation

13. Contract provisions tend to specify wage increments if the rate of change of the consumer price index exceeds some predetermined number. There is often no provision for reducing wage increases for unexpectedly low rates of change of the CPI.

of labor, of course, never matches this ideal. Because the fraction of aggregate demand allocated to any sector follows a random walk, by the specification of equation 1, the extent of excess supply or demand in a sector in any period provides some guide—though an imperfect one—to allocating labor across sectors for future periods, with that information becoming less valuable the further it is extrapolated into the future. Even in the no-contract case, relative wages depend upon past, as well as current, excess supply or demand; hence, wages provide a less efficient guide to labor allocation in this model than does relative excess demand, the term $X_{it} - X_{jt}$ of equation 6. Wage differences will depend on information that is lagged even more with wage contracts than without them. As a result, a somewhat higher NAIRU might be expected in the model with wage contracts than without them.¹⁴ Evidence of this tendency appeared in the simulations: in the no-recession runs the inflation rates for all the contract cases were a little higher than those of the no-contract case shown in column 3 of table 1. It seems likely that in the real world some workers remain unemployed during nonrecessionary periods hoping to regain high-wage jobs they have lost, even though the high wage reflects a contract made under conditions no longer prevailing. If such workers moved, they might reduce the excess demand for labor elsewhere. The impact of wage contracts on the NAIRU is an interesting topic for further research.

Some Aspects of the New Contract Theory

This section reviews the theory of contracts to see how well it explains the existence of contracts in the labor market and the form that they take. It does not pretend to summarize the now extensive literature.¹⁵ While

14. If demand shocks showed negative serial correlation, this would not be true. Wage contracts could provide a smoothing of wages that reduced inefficient labor mobility.

15. The following is only a partial list, roughly in chronological order; and not all the authors may wish to be called contract analysts. Martin Neil Baily, "Wages and Employment under Uncertain Demand," *Review of Economic Studies*, vol. 41 (January 1974), pp. 37–50; Costas Azariadis, "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, vol. 83 (December 1975), pp. 1183–1202; Donald F. Gordon, "A Neo-Classical Theory of Keynesian Unemployment," *Economic Inquiry*, vol. 12 (December 1974), pp. 431–59; Costas Azariadis, "On the Incidence of Unemployment," *Review of Economic Studies*, vol. 43 (February 1976), pp. 115–25; Oliver E. Williamson, Michael L. Wachter, and Jeffrey E. Harris,

the existence and importance of wage contracts has long been recognized, a distinctive feature of the new theory is that it considers implicit as well as explicit contracts. It seeks to explain contract-like arrangements where no explicit legal contracts exist. This is very important because although the observed wage-employment adjustments may be somewhat different in union and nonunion sectors, the general pattern is so pervasive that a "union-only" explanation of wage responsiveness is hard to sustain.¹⁶

The first question, then, is why workers and firms would want a contract that resulted in a path for wages and employment other than that resulting from the instantaneous interaction of prevailing market forces. Both firms and workers have an interest in establishing a long-term relationship—over several years. There is a daily spot market in the shares of General Motors, but not in its employees. For workers, changing employers is costly. Job changes may involve unemployment, search, possibly relocating, and such other costs as giving up friends and accepting many other changes in life patterns. Turnover is costly for firms, too, imposing administrative hiring and firing costs, and loss of specific human capital that workers acquire on the job and of the team effect whereby a group of workers who know each other and the work place perform more efficiently. Thus, a strong element of bilateral monopoly encourages both firms and workers to determine a jointly efficient way to set the wage path and conditions of employment over some extended period. The in-

"Understanding the Employment Relation: The Analysis of Idiosyncratic Exchange," *Bell Journal of Economics*, vol. 6 (Spring 1975), pp. 250–78; Steven Shavell, "Sharing Risks of Deferred Payment," *Journal of Political Economy*, vol. 84 (February 1976), pp. 161–68; Arthur M. Okun, "Inflation: Its Mechanics and Welfare Costs," *BPEA*, 2:1975, pp. 351–90; Herschel I. Grossman, "The Nature of Optimal Labor Contracts" (paper presented at the Third Reisenburg Symposium, 1975; processed); Martin Neil Baily, "On the Theory of Layoffs and Unemployment," *Econometrica*, vol. 45 (September 1977), forthcoming; Martin Feldstein, "Temporary Layoffs in the Theory of Unemployment," *Journal of Political Economy*, vol. 84 (October 1976), pp. 937–57; Robert J. Gordon, "Recent Developments in the Theory of Inflation and Unemployment," *Journal of Monetary Economics*, vol. 2 (April 1976), pp. 185–219 (this paper contains additional references); Herschel I. Grossman, "Risk Shifting and Reliability in Labor Markets," *Scandinavian Journal of Economics*, vol. 79 (1977), forthcoming.

16. As Robert E. Hall has noted in "The Rigidity of Wages and the Persistence of Unemployment," *BPEA*, 2:1975, pp. 301–35, the government and regulated sectors now employ an impressive part of the labor force. One might want to construct a theory of wages for workers in the union plus government plus regulated sectors. Probably this is still not enough, especially as these groups are not entirely distinct.

terest the two parties have in staying together insulates against short-run fluctuations in market conditions.

This kind of argument does not completely explain wage stickiness, however. In practice, the arrangements between firms and workers are *contingent* contracts, under which the income a worker will receive in some future period depends upon the market conditions in that period; and that dependence must be explained. When a firm faces a decline in the demand for its product, it does reduce its labor input promptly and often substantially. On the other hand, it deviates little from the predetermined path of wage rates, especially under an explicit legal contract but also when contracts are only implicit.¹⁷ Even when the wage path is reset or renegotiated in the presence of excess supply of labor, the new path is not chosen to eliminate the excess supply, at least in the short run.

In addition to the behavior of wages, an explanation is required for the choice between shortening the workweek and cutting the work force as a means of varying labor input during recession. Although in practice hours per worker do vary somewhat, the larger part of the adjustment in manhours comes from employment variations,¹⁸ for reasons that require investigation.

In summary, contingent contracts have an obvious appeal for both firms and workers, but the reasons why these contracts result in substantial variations in employment, modest variations in hours, and a sluggish response of the wage path are less obvious.

THE NATURE OF THE OPTIMAL CONTINGENT CONTRACT

One approach to understanding the form of contingent contracts starts with the way a firm and its workers divide up the risks of demand fluctuations. In this view, stockholders can bear risks more easily than workers can because they are better able to diversify their risks. Even though

17. Robert E. Hall has argued that individual wages may be more flexible than scale wages; see "The Process of Inflation in the Labor Market," *BPEA*, 2:1974, pp. 343-93. But allowing for this possibility does not change the basic picture.

18. Arthur M. Okun estimates that about 20 percent of a cyclical addition to or subtraction from labor input was due to hours variations; see "Upward Mobility in a High-pressure Economy," *BPEA*, 1:1973, p. 211. Further analysis of the response of hours and employment can be found in Ray C. Fair, *A Model of Macroeconomic Activity*, vol. 2: *The Empirical Model* (Ballinger, 1976), and Christopher A. Sims, "Output and Labor Input in Manufacturing," *BPEA*, 3:1974, pp. 695-728.

workers may be *able* to borrow and lend—at least, lend—stockholders can operate in capital markets on better terms than workers and, therefore, logically should absorb more of any joint risks.¹⁹ This idea was the basis for a model that showed that firms will actually eliminate wage uncertainty, and surprisingly this result holds even if workers face a positive probability of layoff.²⁰ To understand this consider the following expression, which is the expected utility of a worker in some period:

$$(12) \quad \text{Expected utility} = E \left[\frac{N}{L} U(Y_w) + \left(1 - \frac{N}{L}\right) U(Y_u) \right],$$

where

N = employment

L = the number of workers seeking employment

N/L = the probability of employment

Y_w = the wage income in the period

Y_u = the income of a worker if he is not employed—that is, laid off.

If workers are risk averse, eliminating the uncertainty of Y_w will raise expected utility even if $N < L$ —that is, even if there is some prospect of a layoff. Further, the result holds for *any* positive N , so that it must hold for the optimal N . Of course, in equation 12 one could still trade off a higher probability of employment for greater wage uncertainty and leave expected utility unchanged. But when the firm absorbs the risks, that tradeoff is never efficient. *The employment decision and the wage decision are, in fact, separable.* If two people are partners in a game of bridge, for example, they can follow an optimal strategy to maximize their winnings. This strategy is separate from any decision about the way in which to divide their winnings. A firm and its workers acting in concert will choose the level of production and the employment consequent on that decision in order to maximize their joint welfare (that is, to be on their utility-possibility frontier). The wage then determines how the welfare is divided. This property of efficient contingent contracts is more general than the particular risk-division model described here. It means that the wage and the marginal product of labor are not always equal in the short run, although

19. Moreover, the simple borrowing and lending process does not eliminate uncertainty, which requires insurance or some kind of pooling of risks.

20. See Baily, "Wages and Employment." Among the assumptions used in the model were risk-neutral firms, no borrowing and lending by workers, no variations in hours per worker, and additively separable utility functions.

when hiring a worker a firm will still weigh his expected marginal product against his expected wage cost over the duration of their job attachment. It should be emphasized that reaching the optimal contract will often depend on close cooperation between firms and workers and this in turn depends upon the forces that encourage long-term relationships.²¹

The separability of the labor-utilization decision and the wage decision in an optimal cooperative contingent contract has important consequences. First, somewhat paradoxically it means that even though the risk-allocation model can provide an explanation of the stickiness of the wage path, that stickiness does not *in itself* explain the failure to use the available labor input fully. However, the contract approach suggests looking for an explanation in the opportunity cost of a worker's time. When product demand falls it may be more efficient (privately, but not necessarily socially) to stop using some labor in the production of output. In fact, while risk aversion and such considerations as fairness and wage bargaining are probably important to the stickiness of the wage path, the key issue is the extent to which variations in labor input are derived from a jointly optimal contract.

The central analytical argument takes the following form: If a reduction of labor utilization is simply one of the outcomes of an optimal long-run contingent contract between the firm and its workers, this reduction will not result in excess supply of labor in the usual Walrasian sense and consequently will not exert the same downward pressure on wage rates.

For this argument to be convincing, however, the observed pattern of layoffs and variations in hours must correspond to the pattern that would follow from an optimal contingent contract. Therefore, I decided in subsequent work to focus on the determinants of layoffs and of variations in hours within the context of a contingent contract model, but to drop the assumption of risk aversion. Two independent but somewhat similar models developed by myself and by Martin Feldstein²² derived the re-

21. See the discussion of mobility costs as an important factor sustaining the implicit contract in Baily, "Wages and Employment." The analysis here is focused on the nature of optimal cooperative contracts. But potential conflicts over the division of the joint welfare, and the possibility of default, may lead to contracts that could be improved upon with full cooperation but that are optimal given the costs of negotiation and default. For further discussion of this issue, see also Grossman, "Optimal Labor Contracts."

22. Baily, "Layoffs and Unemployment," and Feldstein, "Temporary Layoffs in the Theory of Unemployment."

sponse to demand fluctuations that is optimal for a firm and its workers combined.²³ The optimal joint response to a fall in demand is to reduce the worker hours of labor input and hence the quantity of output. The division of a reduction of labor input between layoffs and lower hours of work is given by the following equations. For layoffs, the condition is

$$(13) \quad PG'(NH)H - D(H) \geq Y_u - T,$$

and for hours of work the condition is

$$(14) \quad PG'(NH) = D'(H),$$

where

P = the price of the firm's product

N = employment

H = weekly hours of work per worker

$D(H)$ = the disutility of working H hours (related to the utility of leisure)

$G(NH)$ = the production function with capital held constant

$G'(NH)$ = the marginal physical product of an extra hour's work

$G'(NH)H$ = the marginal physical product of an extra worker

Y_u = the income (or income equivalent) received by a worker if he is laid off by the firm—that is, the value of his time if he is laid off (including unemployment insurance benefits received)

T = the turnover cost to the firm of a layoff (including the fraction of the unemployment benefit paid for by the firm through its experience rating).

The intuitive content of these conditions is straightforward. Condition 14 says that hours of work are always set so that the value of the hourly marginal product is equal to the marginal disutility of work. The layoff rule says that only if the left-hand side of 13 is less than the right will layoffs be made, and then in sufficient quantity to restore equality. This, in turn, means that layoffs take place when the net value of working (the

23. To simplify the analysis the firm is assumed to be a price-taker in the product market. Demand fluctuations in the models are price fluctuations. This is not an ideal assumption to use in a model of cyclical demand changes. In fact, it is not consistent with the earlier comments about how changes in aggregate demand influence output. However, the main ideas would carry over to more complex cases. The simple assumption does prevent confusing the issues of labor market response with issues introduced by imperfections in product markets.

value of the worker's marginal product less the disutility of producing it) falls below the net value of a layoff (the value of a worker's time if laid off, Y_u , less the turnover cost, T). Two key parameters in determining the number of layoffs for a given fall in demand are, therefore, Y_u and T . There are two parts to Y_u : unemployment insurance (UI), and the value of time spent searching for and potentially finding an alternative job. The utility of leisure also plays a role in the layoff rule; it is the term $D(H)$ on the left-hand side of equation 13.

In the case of permanent layoffs or discharges, the layoff rule looks sensible. If demand falls for one firm, it becomes worthwhile for some workers to move elsewhere. Equation 13 tells when it is appropriate to make the move, given unemployment insurance to subsidize the transition. But in fact about two-thirds of all layoffs are temporary in the sense that workers return to their former jobs after being recalled by their employers.²⁴ Equation 13 suggests that hours will be reduced substantially before temporary layoffs occur, unless there is a sizable UI benefit that is not effectively experience-rated,²⁵ and implies that a very large fraction of layoffs can be blamed on the UI system. While some evidence suggests that UI does play a role in increasing unemployment and layoffs,²⁶ the historical experience is that extensive layoffs rather than work-sharing formed the basic pattern before the system was established. In particular, UI was a *response* to an existing phenomenon.²⁷ Moreover, the desire of

24. Martin S. Feldstein, "The Importance of Temporary Layoffs: An Empirical Analysis," *BPEA*, 3:1975, pp. 733–36, and "Discussion" (of Thomas F. Bradshaw and Janet L. Scholl, "The Extent of Job Search during Layoff"), *BPEA*, 2:1976, p. 525.

25. Clearly, a statement like this depends upon the interaction of all the parameters. But to support it, consider the following polar case: (1) workers do not search for alternative jobs so that $Y_u = B$, the UI benefit; (2) UI is fully experience-rated so that $T \geq B$ (that is, firms pay the full cost of UI benefits, which in turn are taxable); (3) workers have increasing marginal disutility for work or, equivalently, diminishing marginal utility of leisure: $D'' > 0$. Substituting 14 into 13 yields $D'(H)H - D(H)$ for the left-hand side of 13. This is strictly positive for $H > 0$ by the third condition. Then the first two imply $Y_u - T \leq 0$ so that the three conditions rule out temporary layoffs. This result would stand if the specification of the model were changed by substituting marginal revenue for price, since that would affect equations 13 and 14 equivalently.

26. See, for example, the papers included in the symposium on UI, forthcoming in *Industrial and Labor Relations Review*.

27. Between 1929 and 1932 manufacturing employment fell from 10.7 million to 6.9 million, or by 36 percent. In the same period, average weekly hours fell from 44.2 to 38.3, or 13 percent. There was no UI at that time and other cost factors favoring layoffs, such as fringe benefits and payroll taxes, were very minor.

firms to avoid the loss of firm-specific human capital makes layoffs look even less attractive (in the model this would raise T , the turnover cost).

The key question posed above is whether the theory of the optimal contingent contract provides an empirically valid picture of how labor input is varied. The conclusion seems to be that the model predicts more variations in hours and fewer layoffs in response to a downturn than are actually observed. Invoking the UI system makes the model's prediction more compatible with observation, but this way out does not seem to be satisfactory in the light of historical experience. These tentative results could mean that the basic approach is wrong, but alternatively they could mean merely that the model is incomplete. The latter view is taken in the next section, which seeks to extend the model.

DIFFERENCES IN LAYOFF PROBABILITIES AMONG WORKERS

Most firms have a group of workers who have been with them for a number of years. Although average turnover rates are high, it is striking how many workers actually stay with a single firm for many years. Table 2 gives data on this question from the National Longitudinal Survey of males who were between 45 and 59 in 1966 and who reported in 1966 and in subsequent surveys on the length of their longest spell with the same employer.²⁸ The information in table 2, which applies to completed and uncompleted longest spells as of 1971, reveals that of workers still in their longest spell, 93 percent had been with their employer for ten years or more, and 64 percent for twenty years or more. The percentages are somewhat lower for completed spells; but for the two groups taken together, they are still 86 percent for ten years or more and 52 percent for twenty years or more. Furthermore, these are underestimates, because the first group consists of workers who have not yet completed their longest spell with the same employer.

Along with its long-term employees, the firm will generally also have a group of more recent hires that typically comprises younger workers, but may also include some older workers who have high turnover throughout their working lives. As many people have observed, layoffs are not applied equally to workers in the two groups, but are made in a systematic

28. Survey of Work Experience of Men 45-59 sponsored by the Manpower Administration of the U.S. Department of Labor and conducted by the U.S. Bureau of the Census.

Table 2. Distribution of Male Workers, 45–59 in 1966, by Years of Longest Period with Employer, 1971^a

<i>Employment status and years with employer</i>	<i>Percent of this group</i>	<i>Percent of total</i>
<i>Still with longest employer</i>		
0–4	0.8	0.5
5–9	6.5	3.8
10–19	28.6	16.5
20–29	35.9	20.7
30 and over	28.2	16.3
Total	100.0	57.8
<i>Unemployed or with different employer</i>		
0–4	6.1	2.6
5–9	17.1	7.2
10–19	42.1	17.8
20–29	24.6	10.4
30 and over	10.1	4.3
Total	100.0	42.2
All	...	100.0

Source: Computer tape file, National Longitudinal Survey, U.S. Department of Labor, Manpower Administration. Figures are rounded.

a. Percentages are calculated from sample numbers that have been weighted to reflect the importance of an individual relative to the U.S. population.

way, generally based upon seniority.²⁹ Workers gradually acquire tenure, or at least greater job security, after they have been with the same firm for a number of years. Although there is no really clear-cut line between the two groups, several writers have found these and other differences sharp enough so that they speak of a dual, or “two-tier,” labor market.³⁰

Whatever the reason for the observed pattern of layoffs, it clearly has important consequences for the model of the previous section. In that model every worker was assumed to be treated equally. The tradeoff of layoffs versus reduction in hours was in terms of, say, 10 percent fewer hours or a 10 percent probability of layoff for everyone. But, given the seniority system, the tradeoff for most workers is between 10 percent fewer hours and virtually a zero probability of layoff, while for other workers the layoff probability is close to unity if hours are not reduced. Since layoffs are

29. See, for example, M. W. Reder, “Wage Structure and Structural Unemployment,” *Review of Economic Studies*, vol. 31 (October 1964), pp. 309–22.

30. Peter B. Doeringer and Michael J. Piore, *Internal Labor Markets and Manpower Analysis* (Heath, 1971); Okun, “Upward Mobility.”

generally the major response to a downturn, the burden is distributed very unevenly. Some workers are much worse off while the majority escape with little or no penalty.

There are a number of possible explanations for the layoff pattern, more than one of which fits the contract-theory framework. However, many explanations take layoffs (rather than hours reductions) as given, and merely consider which workers are most likely to be laid off.

COLLUSIVE BEHAVIOR BY WORKERS

The idea that collusion by workers explains the inability of unemployed workers to bid down wages has been around for a long time. The argument made here goes beyond the traditional one in three ways: First, it holds that collusive behavior by workers is potentially much more widespread than unionization. Second, it focuses on the decision to lay off workers who currently have jobs. Third, it remains within the framework of contingent contracts.

Following the approach of the dual labor market, consider a firm with a group (A) of experienced workers, and a group (B) of workers that includes recent hires. If all the group A workers were to quit at the same time, the firm would find it very costly, for it would have to start afresh with completely new workers and would be subject to a team effect, mentioned earlier, losing the familiarity that its experienced workers have with each other and with the operation of the plant. An illustration is the heavy start-up costs of opening a new plant and bringing productivity up to the level of the firm's other plants. To minimize these costs, in fact, firms usually bring in a group of experienced workers from their other plants to help get things started.³¹ These cost factors mean that the group A workers have some monopoly power over the firm. Given this power, a likely reason for the preference for layoffs over work-sharing is precisely the different distributional consequences of the two. The experienced workers wish to protect themselves from the burden of the downturn.

The group pressure to bring about the preferred solution may differ

31. The issues raised here are discussed in Doeringer and Piore, *Internal Labor Markets*. Of course, the capital embodied in individual skilled workers is important. For the most part, however, the returns to this capital are reflected in relative wages. To the extent that firm-specific human capital exists, and the wage differential earned by an experienced worker is less than his marginal-productivity differential, the bargaining power of the group A workers will be enhanced. See, for example, Walter Y. Oi, "Labor as a Quasi-Fixed Factor," *Journal of Political Economy*, vol. 70 (December 1962), pp. 538-55.

in different situations. At one extreme it may be rather subtle, hardly describable as collusion, showing up as “morale problems” or “the workers would not stand for it.” At the other extreme lies a union, which makes collusion explicit. Union members are protected by law against wholesale firing and have the strike as another lever to exert bargaining power.³² However, while unions probably do make a difference, collusive behavior can be much more widespread than unionization. Perhaps unionization is an institutional response to the need to apply group pressure in situations in which the workers do not know each other well enough to do so informally.

Take a period when product demand facing the firm is high. The firm may already have a number of group B recent hires, but it wishes to hire more. Assume that the group A workers have been able to bargain for themselves an expected income that is higher than that necessary to attract workers to the firm but that they are willing to allow an offer to group B workers that is no better than this competitive expected income offer. In fact, to do this is in the interests of the group A workers, since anything that raises the firm’s profits makes it easier for them to achieve a higher expected income for themselves.

Since the firm can utilize group B workers and need offer them only the competitive expected income, it will indeed do so. Group A and group B workers then will have different contracts. But what form will they take and in particular, how will they differ? A fully cooperative optimal contract between the firm and its group A workers, that maximized the firm’s profit subject to a given expected income of group A workers and subject to the market constraint that group B workers must receive the competitive expected income, would have the following properties: (1) group B workers would receive a lower wage than group A workers; (2) the breakdown between variations in hours and layoffs would be the same as in the model of the previous section.³³ However, such a contract

32. Nonunionized workers also have means to pressure the firm. Morale problems translate into productivity problems. There can be go-slows or working-to-rule without a union.

33. If both types of workers have the same disutility-of-work function, $D(H)$, hours of work will be the same for both. Otherwise $D'(H^A) \neq D'(H^B)$, which cannot be optimal. Layoffs would commence at the same point as that given by equation 13. If not, it would be possible for group B workers to make side payments to allow them to continue working. Such a possibility (even if side payments are not actually feasible) shows that any rule other than 13 is not optimal. The difference in expected income between the two groups results from paying a higher hourly wage to the group A workers.

will *not* be accepted by the group A workers because of their quite realistic fear of default by the firm. Workers do not in general know all aspects of the situation facing the firm; in particular, they cannot observe the extent of falls in demand, especially when such falls do not show up fully as declines in price. Nor can they determine the turnover cost, which includes the cost of UI benefits paid for by experience-rating. As long as there is no gap between the expected income offered by the firm and the competitive expected income offer, this lack of information is not so vital. A firm that cheats its workers will start to experience difficulty in hiring as its bad reputation spreads. When there is a gap between the expected income group A workers have bargained for themselves and the market expected income, cheating or default can occur without creating difficulties in hiring.

Group A workers must restrict the firm either by specifying the amount by which hours can be reduced, or by forcing the firm to pay the same wage (adjusted for productivity) to group A and group B workers. Otherwise, the firm can easily default on the expected-income bargain obtained by the group A workers.³⁴ All it need do is hire more group B workers in the boom than would be justified by the cooperative optimal contract, and then reduce hours across the board more in any downturn than would be optimal under that contract.³⁵

If there is a direct restriction on reductions in hours (no short time, for example), which certainly will be combined with a provision that puts secondary workers first in line for layoffs, the observed preference for layoffs is clearly understandable. In that event, group A workers need not specify the wage paid to group B workers. The firm will set that wage just high enough to compensate for their layoff probability. If, instead, the group A workers require that the wage paid the two groups be the same, the firm no longer has any incentive to default by substituting cheaper group B hours for group A hours; that is, given a wage restriction, the group A workers need not restrict variations in hours. A simple explicit model with the two types of labor receiving the same hourly wage, but

34. The argument made here is analogous to the one made by Okun in "Inflation," p. 369, concerning "clean hands" policies.

35. Such a strategy effectively substitutes the cheaper worker hours of group B for the more expensive worker hours of group A. If a firm has eighty group A workers earning \$6 per hour working 41 hours per week, it can lower costs by hiring two group B workers at \$4 per hour and reducing everyone's weekly hours to 40. Total manhours are the same. The average hourly wage has fallen.

different levels of expected income, has the following properties. Instead of a layoff rule of the form of equation 13, there is a layoff rule for group B workers of the form

$$(15) \quad PG'[(N^A + N^B) H] \geq f(Y_{u,i}, T, H, V^A - V^B), \\ f_1 > 0, f_2 < 0, f_3 < 0, f_4 > 0,$$

where f is a relation depending on the shadow prices of the two types of workers; and V^A and V^B represent the expected incomes of the two types of worker with $V^A > V^B$ by assumption. The condition that sets hours of work has the same form as 14:

$$(16) \quad PG'[(N^A + N^B) H] = D'(H).$$

Equations 15 and 16 imply that, in the event of a downturn (1) there will be some reduction in hours worked; (2) there will *always* be layoffs of group B workers; (3) once the pool of group B workers is exhausted, the model will behave like the model of the previous section; (4) UI that is not fully experienced-rated will raise the number of layoffs for a given fall in demand.

This analysis does not pin down whether work-sharing will be restricted by collusive workers directly, or indirectly by restricting wage differentials. In either case, the basic amendment to the earlier model results from incorporating the distributional consequences of work-sharing versus layoffs when a group of workers has the power to affect the distribution. The relevant contingent contract, explicit or implicit, is between a firm and a subset of its workers.

FACTORS MITIGATING "PURE" COLLUSION

A key element of the collusive model was the ability of group A workers to raise their expected income above the level needed to attract new workers. Without this feature it is not possible to impose the high layoff probability on the group B workers and still attract them to the firm. However, the group B workers might take a longer view, recognizing their upward mobility as group A workers retire or quit. Because of the prospects of moving into the preferred group, a new hire may accept a job with a short-run expected income actually lower than he or she could obtain elsewhere.

At its limit, this immediately preceding argument asserts that concentrating the burden of the reduction in labor input on a few junior workers

is simply an optimal allocation of the risks of demand fluctuations across different age groups. Although the usual specifications of life-cycle utility functions do not suggest that people will become more risk averse as they grow older, both introspection and casual empiricism uncover the strong desire of many people for increasing income security over their lifetimes.³⁶ During some periods in the life cycle children and mortgages make workers very anxious to avoid income fluctuations.

In addition, if risk aversion is to be put back into the analysis, the risk model described earlier has an important corollary: the greater is the ratio of wage income to unemployment income, the more complete is the implicit insurance policy a worker would wish to purchase to reduce his probability of income fluctuations.³⁷ The senior skilled workers have higher pay and are likely to have a higher ratio of wage income to unemployment income.³⁸

These factors, however, do not provide a satisfactory alternative to collusion as an explanation for the predominance of layoffs over reductions in hours. High-wage or very risk-averse workers may indeed wish to pay an insurance premium to reduce uncertainty, but the optimal insurance policy would be to reduce the correlation between the wage income and hours of work. In the limit the worker would become salaried. Thus, these arguments might be invoked to explain the pattern of implicit contracts for white-collar workers, but they are much less convincing for blue-collar workers.

I conclude, therefore, that collusion retains a role in the explanation of the response of the labor market to recession. The prospect of upward mobility, however, will certainly mitigate the consequences of collusive behavior on the supply of group B workers to the firm. It will also ration-

36. The difficulty with asking such a question is that people tend to assume one is asking "*given the present*, do I want the future to be better?" The answer is always yes. To a young person, a job that implied a constant risk over his lifetime is not obviously less desirable than a job that concentrated the instability now—witness the fact that people often save very little to provide for income security in retirement.

37. This point is analyzed by Azariadis, "Incidence of Unemployment." He does not consider hours variations, however, and possible differences in Y_u among workers.

38. Although not necessarily so. In Massachusetts, for example, the UI benefit is 50 percent of the wage up to a limit, so the ratio is the same for many workers. Many workers first in line for layoffs are not eligible for UI benefits. Further, workers aged 50–60 may have a greater disutility of work than workers 20–30.

alize the distribution of layoffs in a political sense, making it more acceptable and making it seem fair—even though it may sometimes be rather unfair, especially to those workers who remain chronically in group B.

I conclude, further, that the predictions of the optimal contingent model, modified to take into account the distributional consequences of layoffs, are consistent with observations on the response of the labor market to demand fluctuations.

A Request-Control System

The theoretical analysis of contracts has implications for wage setting that introduce inflationary momentum into the process. Any reduction in the demand for labor that is perceived as temporary will not change the terms of the implicit or explicit contract between a firm and its workers. The wage is set to track a long-run equilibrium path or a permanent wage path. This long-run path depends on market conditions, but only on those that are also seen as long run. Because of aversion to uncertainty, because of possible institutional difficulties in revising complex wage scales, and, not least, because the wage represents the result of bilateral bargaining, setting the permanent wage path in practice usually means setting wage rates for periods of one, two, or three years. Wage contracts, like other contracts, are set in dollars or in nominal terms, but, like other contracts, they will embody information about other variables. These will include wages elsewhere, the firm's product price, and the cost of living. The actual movements of these variables over the period of the previous contract will be compared to the movements that had been anticipated, and their future movements will be predicted. Thus, the momentum of inflation gets built into the path of the nominal wage.

Even when the wage is renegotiated the wage path will be adjusted *only to the extent that the perception of the permanent wage has changed*. Hence, a contract, implicit or explicit, that is reset during a recession will not be reset generally in such a way as to eliminate any current excess supply of labor in that sector. The importance of long-run considerations is consistent with theory and observation and is embodied in the simulation model. This long-run focus may be particularly strong when firms and workers realize that the policymakers are likely to restore the NAIRU after a few years.

The analysis above has stressed that the existence of wage contracts—the fact that wages are *set*—may be a key factor in explaining the difficulty of slowing an established inflation by means of a recession. Space does not permit examining contracts in product markets, although they probably exist, in both explicit and implicit forms. Some may fit the pattern of customer markets, as Okun has analyzed them;³⁹ but others probably involve the implicit collusive relation among sellers. The important aspects of pricing for this paper are the following: (1) Prices or price markups do not respond enough to alleviate the problem of inflation control that has been described in the labor market. (2) Most prices as well as most wages are *set*. Firms are not price-takers, at least over the short run, and this introduces an important common element on the price and wage sides. Both prices and wages are set and are changed only periodically.

The basic dilemma in using wage and price controls as an anti-inflationary policy is that, unless a great deal of information about market conditions in thousands of markets is available so that market-clearing relative wages and prices are known, the controls impose distortions of relative prices and wages. Of course, the model of a market economy that implies (Pareto) optimal conditions is, in fact, so far removed from the real market economy that one wonders how costly the supposed distortions really are. In particular, the very fact that the real economy responds as it does to recession is evidence that the assumptions required for allocative optimality are not satisfied.

Nevertheless, in an economy based on the market system the imposition of controls in the absence of a theoretical basis that describes how they should work, if only in principle, is a step taken with great reluctance. This section presents a theoretical outline for a system of nondistorting wage and price controls. The system described below is not intended to be immediately practical, or to be applied directly to the economy. However, just as knowing the conditions under which a market economy would be Pareto optimal is a helpful guide and reminder for policymaking, knowing the conditions under which wage and price controls would be nondistortionary may be helpful in their actual operation.

39. Okun, "Inflation."

A THEORETICAL OUTLINE OF THE SYSTEM

The model economy in which the “request-control” system is considered is characterized by wages and prices that are set and held over some time period; it is a contract economy. In this economy n products are produced. The output of the i th product is Q_i and $i = 1, \dots, n$. The division into products is unorthodox: The output of each firm in the economy is considered to be a separate product and, for simplicity, each firm is assumed to produce only one product; thus, n is just equal to the number of firms. The set of products will be divided into three groups: the first m are sold to final use; the next k are capital goods; and the last q are intermediate goods. Thus, $n = m + k + q$.

In the absence of any controls, each firm is assumed to set the price of its product and to maintain it for one period. In period t , therefore, this gives a price vector (\bar{P}):

$$(17) \quad \bar{P}(t) = \{P_1(t), \dots, P_n(t)\}.$$

Each firm is assumed also to set the wage it wishes to pay in period t , with or without explicit bargaining with the workers. This gives a vector of wages (\bar{W}):

$$(18) \quad \bar{W}(t) = \{W_1(t), \dots, W_n(t)\}.$$

This formulation implies homogeneity of workers, although it could be extended to heterogeneous workers.

It is assumed that wages and prices are set for the same period and are not varied within it. In periods $t + 1$, $t + 2$, $t + 3$, and so on, new vectors of wages and prices are set. Since the application of controls is going to be considered, it is assumed that the economy is following an inflationary path of rising wages and prices.

Using the vector of prices set in any period t , it is possible to construct a price index ($I(t)$). Abstracting from the practical difficulties of constructing an appropriate index, I assumed weights $\lambda_1(t), \dots, \lambda_n(t)$ ⁴⁰ such that the rate of change of I is the measure of inflation relevant for eco-

40. The argument t implies that the weights are *used* in t . They may be based on prior quantities. Many of the λ s are presumably zero.

nommic policy; and its level is the appropriate price deflator of wage and profit income to compute real wages and profits.

$$(19) \quad I(t) = \frac{\sum_1^n P_i(t)\lambda_i(t)}{\sum_1^n P_i(0)\lambda_i(t)};$$

The monetary and fiscal policy applied to the economy is such that, given any set of initial conditions, it will generate time paths for the output of the n goods, the labor and capital inputs used to produce them, and the vectors of prices and wages. The request-control system to reduce the rate of inflation in the model economy clearly must have some real consequences (if not, it has no point). However, conditions will be specified under which the controls cause minimal distortion.

At the beginning of each period, each firm must submit requests for the price it wishes to charge for its product (P_i^r) and the wage it wishes to pay its labor. The price requests can have any magnitude and workers can continue to bargain in setting the firm's wage requests. Based upon the vector of price requests submitted, the price-request index, $I^r(t)$, is

$$(20) \quad I^r(t) = \frac{\sum_1^n P_i^r(t)\lambda_i(t)}{\sum_1^n P_i(0)\lambda_i(t)}.$$

This expression corresponds to the price level that would obtain if the requests became the actual prices during period t .

Suppose policymakers have determined a policy goal for the level of the price index for t , which will be denoted $I^d(t)$. That goal and equation 20 define a scaling factor $\alpha(t)$, which is independent of i .

$$(21) \quad \alpha(t) = I^d(t)/I^r(t).$$

Vectors of controlled prices and of controlled wages are defined by

$$(22) \quad \begin{aligned} P_i^c(t) &= \alpha(t)P_i^r(t), \quad i = 1, \dots, n \\ W_i^c(t) &= \alpha(t)W_i^r(t), \quad i = 1, \dots, n. \end{aligned}$$

The price $P_i^c(t)$ and wage $W_i^c(t)$ are then sent back to all firms $i = 1, \dots, n$, to serve as their controlled levels. The period that was previously a contract period is now a control period; and firms and workers are required to set wages and prices that are *neither higher nor lower*

than the control levels for the duration of a control period. At the beginning of each new control period, $t + 1, t + 2, t + 3, \dots$, the process is repeated, aimed at a policy-determined path of the desired price level that reflects the target rate of inflation. New vectors of price and wage requests and new vectors of controlled prices and wages are set each period and, of course, a series of scaling factors $\alpha(t + 1), \alpha(t + 2)$, and so on.

Provided the wage and price requests are all positive and finite, and that the controls are enforced, the actual price level will equal the desired price level in any period for which the controls are applied:

$$(23) \quad I^c(t) = \frac{\sum P_i^c(t)\lambda_i(t)}{\sum P_i(0)\lambda_i(t)} = \alpha(t) \frac{\sum P_i^r(t)\lambda_i(t)}{\sum P_i(0)\lambda_i(t)} = I^d(t).$$

It can now be seen how the role of worker bargaining for wage requests is maintained, as was stated above: The greater the gains the workers obtain in the request, the greater the wage they finally obtain in the control.

The characteristic that the controlled prices (and wages) be binding *both upward and downward*—hence they are *not* merely price or wage *ceilings*—is a key, distinctive, feature of the system and is essential to the properties discussed below.⁴¹

The most important property of the controlled wages and prices is that the relative wages and prices and the real wages imposed by controls are equal to the relative wage and price and the real wage requests submitted. This property, from equations 22 and 23, holds for any strictly positive value of $\alpha(t)$, the scaling factor. Thus,

$$(24) \quad \begin{aligned} (a) \quad & P_i^c(t)/P_j^c(t) = P_i^r(t)/P_j^r(t) \text{ for all } i, j \text{ and any } \alpha(t) > 0 \\ (b) \quad & W_i^c(t)/W_j^c(t) = W_i^r(t)/W_j^r(t) \text{ for all } i, j, \text{ and any } \alpha(t) > 0 \\ (c) \quad & W_i^c(t)/I^c(t) = W_i^r(t)/I^r(t) \text{ for all } i \text{ and any } \alpha(t) > 0; \end{aligned}$$

One difficulty with applying controls of the usual sort is the huge volume of information required to set nondistorting wages and prices. The re-

41. Some suggest that it would be hard to gain acceptance for a regulation against lowering prices (over the control period) and hard to enforce it. However, laws against cutting prices already exist in a number of markets, particularly in transportation. While they are widely considered undesirable, they do suggest that such laws are politically feasible—and might even be feasible when they serve a more desirable purpose. Enforcement would be aided by the incentive rivals would have to complain about a firm that was cheating on its controlled price during the control period. But the request system need not discourage price-cutting in general, since firms can always submit low price requests.

quest-control system solves this problem: by setting the controls according to equation 22 and enforcing them in both directions, the system extracts from firms and workers the information the controllers need about *relative* wages and prices. Whatever firms request as a relative price or a relative wage they get when the controls are applied.

Making the assumption (discussed further below) that the real interest rate $R(t)$ can be made independent of the scaling factor $\alpha(t)$ yields the next property of the system: the real rental rate on capital is independent of $\alpha(t)$ also. To demonstrate this proposition, define $G_h^e(t)$ as the expected rate of change of the ratio P_h/I for $h = m + 1, \dots, m + k$ (where h is a capital good). Hence $G_h^e(t)$ is the expected real capital gain or loss from holding capital good h . Now just as equation 24c followed from equations 21 and 22, so also does it follow that

$$(25) \quad \frac{P_h^e(t)}{I^e(t)} = \frac{P_h^i(t)}{I^i(t)} \text{ for } h = m + 1, \dots, m + k \text{ for any } \alpha(t) > 0.$$

As was noted, the control system is assumed to be applied in periods subsequent to t in the same way as specified in equations 21 and 22. If by setting $\alpha(t) = 1$ for all t greater than some T the controls can actually be removed without changing equation 25, it follows that $G_h^e(t)$ is independent of $\alpha(t)$ and the path of α in the future. The nominal rental rate on capital good h (C_h) is then given by

$$(26) \quad C_h(t) = P_h(t) [R(t) + D_h(t) + G_h^e(t)].$$

And since R , G_h^e , and clearly D_h (the rate of physical depreciation) are all independent of $\alpha(t)$, it follows that

$$(27) \quad \frac{C_h^e(t)}{I^e(t)} = \frac{P_h^e(t)}{I^e(t)} \{R + D_h + G_h^e\} = \frac{C_h^i(t)}{I^i(t)}.$$

The real rental rate on capital good h for $h = m + 1, \dots, m + k$ is the same as that implied by the requested prices—*independent of $\alpha(t)$* .

The next important property of the system is that the quantity of output of the i th good, Q_i , and the quantity of labor employed to produce this output, N_i , that maximize real profits, π_i/I , are the same for the controlled prices and wages as for the requests—*independent of $\alpha(t)$* .

$$(28) \quad \frac{\pi_i^e}{I^e} = \frac{1}{I^e} \{P_i^e Q_i - \bar{P}_i^e \bar{Q}_{i_q} - W_i^e N_i - \bar{C}_i^e \bar{Q}_{i_k}\}$$

$$\begin{aligned}
 &= \frac{1}{I^r} \{ P_i^r Q_i - \bar{P}_q^r \bar{Q}_{iq} - W_i^r N_i - \bar{C}_k^r \bar{Q}_{ik} \} \\
 &= \frac{\pi_i^r}{I^r} \text{ for all } i, \text{ for any } \alpha(t) > 0,
 \end{aligned}$$

where \bar{Q}_{iq} is the vector of intermediate goods purchased by firm i , \bar{Q}_{ik} is the vector of capital goods rented, and \bar{P}_q and \bar{C}_k are the corresponding vectors of prices and rental rates. In order for this result to be correct, real demand for all firms must also be independent of $\alpha(t)$. This is discussed further below.

Apart from the assumptions about the interest rate and demand conditions, the main implications of the system are now clear. Real wages, real profits, relative prices, and relative wages are all independent of the $\alpha(t)$ chosen—in particular, they will be the same for both $\alpha(t) < 1$ and $\alpha(t) = 1$, the case of no controls; it follows that the price and wage requests submitted should be equal to the prices and wages that would have been set in the absence of controls; and that the path of real output and employment in all sectors should be unchanged by the imposition of controls.⁴² These conclusions are a natural extension of the well-known neutrality of a general-equilibrium system to the price level. It was, perhaps, not obvious a priori that a system of requests and controls could exploit this property.

Finally, the important assumptions about the real interest rate and real aggregate demand require inspection. Suppose the economy is, in the absence of controls, operating at the NAIRU with a positive rate of inflation. For simplicity, suppose that only monetary policy can influence the level of aggregate demand. Therefore, in the specified steady state the growth of the money supply is just sufficient to accommodate the growth of the real output necessary to maintain the NAIRU and the positive rate of inflation. In period t the controls are imposed with $\alpha(t) < 1$, so that the rate of inflation is to be reduced. What determines the accompanying monetary policy? Let the money stock in t be $\alpha(t)$ times the money stock in the absence of controls. Then the rate of growth of the money supply can be set just to accommodate the growth of real output

42. One of the assumptions underlying this result is that firms synchronously set wages and prices at the beginning of each contract period even in the absence of controls. Of course, this is not true, and the implications of nonsynchronization are discussed in the next section.

plus the rate of inflation along the path of controlled prices. This initial adjustment in the money supply and resetting of its growth rate would approximate the required conditions. However, to the extent that imposing the controls and slowing money growth would lower the expected rate of inflation in future periods, the demand for money and other financial assets would be raised and a slightly higher real rate of interest would be required for the maintenance of equilibrium in the asset market. This should be no surprise since the reduction in the demand for money caused by expected inflation is emphasized as a major cost of inflation, and thus in practice a rise in the real interest rate is desirable. But the current discussion is not about the ideal policy; it is aimed at finding conditions under which the imposition of the controls would cause the minimum change in the paths of real variables in the economy.

The solution is to invoke fiscal policy in order to affect relative prices. Taxation clearly influences the calculation of the effective real rate of interest for individual investors.⁴³ With a delicate hand in setting tax rates, therefore, it is possible to devise a combination of monetary and fiscal policy that allows the real interest rate, the real money stock, and hence (under the simplifying monetarist assumption) the path of real aggregate demand to be independent of the path of $\alpha(t)$.

In summary, the request-control system need cause little distortion in this simplified economy. Real wages, real profits, relative prices, the real interest rate, and real aggregate demand could be untouched by the controls. Since aggregate demand and relative prices are unchanged, the distribution of demand by sector will also be unchanged, as was assumed earlier.

Of course, in common with any anti-inflationary policy, the request-control system, by shifting the price level and its rate of change, would affect debtor-creditor relationships and the structure of interest rates. It might also exert a wealth effect because of the change in the real value of government bonds. These impacts have been ignored; but it is not clear that they should be considered distortions in any case, nor do they change the fundamental property that the system allows relative prices to adjust while controlling the rate of inflation.

43. Robert E. Hall and Dale W. Jorgensen, "Tax Policy and Investment Behavior," *American Economic Review*, vol. 57 (June 1967), pp. 391-414.

CONTROLS: FURTHER ANALYSIS

The pure theory of controls outlined above rests on many strongly simplifying assumptions. To point out directions for further analysis, this section addresses some of the ways in which the controls might function in a more realistic setting.

In the context of the request-control system, the most important characteristic of the contract economy is that wages and prices are set. In practice there are many price-followers, small firms that are not exactly competitive price-takers but whose prices and wages are the same as, or set in relation to, those of the large firms in their industries. There are also a number of auction markets, generally markets for homogeneous raw materials or foodstuffs, where prices are set by supply and demand. And some prices, notably the OPEC oil price, are exogenous for the United States, even though they are not competitive auction prices. Therefore, the request-control system would operate in a nonhomogeneous economy and, as a result, would not be applied to all prices and wages, but only to the wages and the value-added prices of large firms.⁴⁴ The demonstration of the nondistorting properties of the request-control system breaks down when not all prices are controlled. The key issue is the extent to which the combination of controls on the contract sector and a monetary-fiscal policy consistent with the desired rate of inflation would keep other prices in line with the controlled prices.

Since the controlled firms would buy raw materials—and possibly intermediate products—whose prices are not controlled, it is an important feature of the system that it controls value-added prices. Variations in prices of raw materials would be required to be passed through to prices of final goods. In fact, since wages are controlled separately, the firms could simply be required to submit wage and markup requests, which in the event would amount to controlling the value-added price.⁴⁵

44. The dividing line would not, therefore, be set by product. Such a division leads to absurdities: unprocessed corn is not controlled, but slightly processed corn is. Drawing a dividing line according to size is not easy either, of course; presumably it would give some advantage to small firms. I do not find this a compelling problem.

45. Also, the requests would cover the average markup across a product line and the average wage across employees or classes of employees, rather than each product of a multiproduct firm and each wage rate.

The controls could be enforced by a system, like customs checking, of reviewing the accounts of a sample of firms. All firms would be required to keep accounts avail-

Another important characteristic of the pure version of the control system is that all firms were assumed to set new wages and new prices at the beginning of each period. The control system then simply required that these wages and prices be submitted as requests. In practice, of course, not all wages and prices are set at the same time. This issue could be dealt with in part by enforcing greater synchronization of price-setting and in part by reducing synchronization within the control system—for example, by breaking it down into wage requests and controls, and markup requests and controls.

No serious difficulty would be encountered in requiring all large firms to submit requests for their average markup every period (for whatever length is decided—say, one or two quarters). A firm would not have to change its request from one period to the next. Further, specifying that the requests all be made at the same time does *not* mean that the magnitude of a firm's request is kept secret up to the last second. To do this would generate a great deal of uncertainty, especially in industries marked by price leadership. Probably, certain key firms would announce their requests slightly in advance of the submission date. Other firms would follow this lead, just as they now do in pricing, and all the submissions would then be made together. Uncertainty is also reduced by requiring markup requests. Unexpected increases in the costs of inputs would not cause sudden losses, for the increased costs would be passed through to final prices.

The heart of any system of controls is its ability to hold down wage increases. Here the problem posed by synchronization is potentially more severe, but a number of procedures are available to solve it. One involves some compromise with the theoretical control system. At the beginning of each period only firms whose wages are scheduled to rise during that period are required to submit wage requests.⁴⁶ These requests are then scaled down and sent back as wage controls. Other firms simply continue to pay the wages they have been paying. Since there will always be a large number of firms submitting requests in any period, the scaling factor, α , is independent of any single firm's request. However, under this compro-

able in a certain form. Much of the required accounting is work that is already done, either for tax or internal-policy purposes. A firm would calculate its average markup over all costs.

46. The period would probably be a quarter. However, the procedure presumes that most firms set annual increases—that is, one-fourth of the controlled firms submit requests in any quarter.

mise system relative wages are subject to change. But it should be remembered that any anti-inflationary policy will cause a similar distortion. As was observed in discussing the simulation model, wages set by multiperiod contracts embody expectations of inflation that are then invalidated by an anti-inflationary policy.

The final issue is the extent to which the system would encourage firms and workers to raise their price and wage requests above the levels they would have set in the absence of controls. The analysis of the previous section says they would not do so at all. The basic reason, it should be remembered, is that in the absence of controls, firms and workers already set wages and prices, particularly wages, anticipating that inflation will effectively scale them down. Even so, one might argue that firms and workers would perceive the controls as something different which they must try to outwit. As long as only a few try this strategy, the penalty of the market will work to discourage it, for though they will end up with wages or prices above their equilibrium relative wages or prices, they will experience lower quantities demanded over the control period. If all firms and workers try the strategy, the desired path of inflation could be maintained simply by reducing the scaling factor. The only serious potential problem this antidote could pose would be an increase in inflation—a bounce-back—after the controls were removed. But four factors suggest that this need not occur with the request-control system. First, monetary-fiscal policy would have been coordinated to make it consistent with the rate of inflation tracked by controls. Second, since there are no pent-up distortions of relative prices, one of the usual motivations for a post-control spurt of inflation is removed. Third, the controls can be gradually, rather than suddenly, removed simply by letting the scaling factor tend to unity. Fourth, since the scaling down would end after the inflation rate has been low and steady for a number of periods, it is hard to see why the economy would suddenly jump off the steady-state path it was then following.

Concluding Remarks

This paper has examined some implications of contracts, particularly wage contracts, for macroeconomic policy. It used a simulation model to show how the reduced frequency of wage setting associated with con-

tracts may make recession less effective in slowing inflation. Contract theory was extended to explain, first, the existence of wage contracts, implicit or explicit, and, second, the failure to set contract wages so as to eliminate short-run excess supply of labor. According to the argument here, when a reduction in labor use is simply the (privately) optimal response of a firm and its workers to demand fluctuations, labor is not in excess supply in the usual sense. At first inspection the theory behind this explanation seems to predict more work-sharing than is observed. A pattern of collusive behavior by workers was invoked as an explanation of the predominance of layoffs over work-sharing.

Finally, it was argued that in a contract economy of wage and price setting, wage and price controls might be suitably applied. A request-control system was explored, first at a very theoretical level and then subject to some of the modifications necessitated by the complexity of the real economy.

Comments and Discussion

William D. Nordhaus: Martin Baily has prepared a useful and interesting examination of contract theory and its implications for inflation theory. My comments will review three aspects: (1) the general theoretical questions raised in contract theory; (2) the simulation of the differential effects of contracts in recession; and (3) the price-control system.

First, contract theory has performed the very useful function of making more rigorous rather vague notions about how employers are constrained by “morale” or “informal understandings.” These vague notions have been invoked whenever the conundrum of sticky wages or prices is raised. There is obvious contract theory and deep contract theory. Obvious contract theory simply notes that there are long-term contracts that set some prices and wages. Much of the short-run inflexibility of prices and wages, and the momentum effects often discovered, simply reflects this fact. Deep contract theory, as presented by Baily, is more subtle. Its basic notion is that contracts are written over what might be called the economic climate rather than the economic weather: agreements take into account mainly the distributions of economic variability rather than their realizations. As long as a recession is just a rainy day—and does not increase people’s subjective probability that rainy economic weather will be more prevalent in the future—the long-run relative valuation of working and swimming will not change.

Although I believe there is some verisimilitude in the phenomena being described, I have some reservations about the theory insofar as it is deep or “implicit” rather than obvious or “explicit.” If in fact the contracts are not legally enforceable, why are there not the usual powerful incentives to shade them on the part of the individual firms? Why is “contract shading”—cheating a little on the implicit contract—any more implaus-

ible than wage or price shading? Is the rationale that people don't shade wages because of implicit contracts somewhat circular for this reason?

Another set of problems with the contractual theories is that they do not answer the fundamental question: why do explicit contracts cover three years for labor in the United States, a lifetime for labor in Japan, and zero for auction goods? Why is oil a customer good, while copper is not? I have a suspicion that the contract theorists focus on the wrong phenomenon—or at least overemphasize the point—when they argue that the purpose of the contract is to reduce uncertainty, reflecting worker risk aversion vis-à-vis wage changes. It seems more probable to me that the existence or length of a contract depends just as much on the transactions costs of arriving at a bargain, of spelling out all the contingencies in the contingent contract. If the fundamental reason for long-term contracts is to reduce these transactions costs, many of the implications that Baily draws for the cyclical behavior of wages and employment are less plainly applicable. The cyclical sensitivity of wages and prices may lessen, as in obvious contract theory, simply because the response is slower. Whether any additional long-run stickiness is attributable to this source is unclear. I personally am not convinced that the differential diagnosis of implicit contract theory for explaining the stickiness of wages and prices or the persistence of unemployment is proven.

Next, I wish to discuss the simulation exercise in the first part of the paper. Baily has asked a number of interesting questions about the operation of Tobin's labor-market simulation model. The model shows a natural rate of unemployment (or a NAIRU) in that it is inflation-neutral—all real variables will operate at the same level for any given rate of inflation to which the system has completely adapted. Two results emerging from the model were at least a little surprising to me. First, the natural rate appears to be less constant than the speed of light—although perhaps not so whimsical as the velocity of money over the last couple of years. It is clearly quite different with different contract specifications, with a lower natural rate associated with the very quick adjustment of no contracts, and a higher natural rate associated with the very slow adjustment of long contracts. I presume that longer contracts would provide more smoothing, thereby removing some of the inflationary influence of the nonlinearities in the system, and lead to a lower natural rate. This is not so in Baily's simulation. Why? Because the wage rate responds to the current unemployment rate no matter how long the contract period, thus

amplifying, rather than averaging, the noise. Presumably, if the noise is at the red rather than the violet end of the spectrum, long contracts should do more smoothing, taking an average of past unemployment rates in setting wage demands. From a couple of simple scratchings, I have not convinced myself whether or not this would reverse the Baily result.

The simulation results in table 1 also show two effects, one not surprising, and one terribly surprising. The unsurprising result is that the short-run response of inflation to unemployment is much smaller with the friction of contracts imposed upon it. The surprising point is that the long-run response is also smaller in the contract system: put differently, long contracts delay the response not only for a while but forever.

This is not obvious. It is easy to see why it cannot be due simply to the existence of a longer lag of wages behind unemployment when there are price responses. In a simple linear system in which contracts simply lag excess demand by a fixed period, wages are passed through into prices:

$$\dot{W}_t = \alpha X_{t-\theta} + \dot{W}_{t-1},$$

where a period is the time it takes for prices to respond to wages, and θ is the average contract period. Clearly, then,

$$\Delta \dot{W}_t = \alpha X_{t-\theta}.$$

Thus the length of the lag shows up not in the long-run response of wages to demand, but only in the speed of adjustment.

Why does the Baily simulation not show the same properties? He ran the simulation for 120 periods, and doing so did not change the shape of the results. Nor do nonlinearities. Pretty clearly, the wage equation has changed in a subtle way: In programming the change in wages for a multiperiod contract, the wage deceleration arising from excess supply takes place only in the first period. Thus, if the theory is, say, that a one-point increase in unemployment over the natural rate for one year lowers the rate of wage inflation by one point, I would have guessed that wage deceleration would continue over the length of multiperiod contracts; but Baily applies it only to the first period.

A simple example will make this clear. Assume that there is a one-period recession in period 1, all wages rising at 10 percent annually. In a noncontract world, suppose that all wages and therefore the average will rise at 9 percent annually after the recession. Next consider a case of two industries, with nonsimultaneous two-year contracts. Only in the

industry that renegotiates in period 1 will wages be affected in that period. I would have thought that the logic would call for its wages to decelerate at 1 percent per annum *for both periods*, or to rise by 9, and then 8 percent annually over the two years of the contract. In that event, the long run is exactly the same as it is in the no-contract system. In Baily's simulation, however, the inflationary process in the contract world never catches up because the full contract or "life cycle" response of wages, measured as the deceleration in percent per annum for a given change in unemployment, is assumed to be inversely proportional to the contract length. As this is not especially plausible, I take the results to be an interesting feature of the model, but not necessarily an interesting feature of the real world.

One intriguing issue that is raised but not pursued involves rational expectations. It would be very useful to know whether perfect foresight about the systematic noise in the economy (either the persistence in excess demand in a particular industry or the prevailing macroeconomic westerlies) would lead to a markedly different structure. To the extent that local turbulence is more important than the prevailing winds, I suspect that the results would not be a twig-like Keynesian structure easily blown down by the big bad monetarist wolf.

Finally, I would like to address the "request-control system." It seems to me that Baily has set forth a system of price controls obeying the fundamental axiom of a well-designed incomes policy: that to be politically successful, it must be distributionally neutral. An incomes policy is like an attempt to get everybody at a football game to sit down at an exciting moment—when the situation is overheated, so to speak: if everybody sits down people can see *better* and will be a lot more comfortable. The elegance of Baily's scheme is that inflation is reduced by making everyone sit down—not just one's least favorite steel company or labor union.

This having been said, it is also clear that such a scheme is both theoretically and practically impossible. It is theoretically impossible because there is always one nominal price that cannot be reduced—the nominal rate on money. On a practical level, Baily himself notes many of the problems. The basic difficulty is that *the scheme forces firms to announce relative prices* and, what is more important, to stick to them. This *will* work in the Arrow-Debreu world of perfect futures and insurance markets clearing now for all time to come (although why inflation is a serious problem in such a world is a bit of a puzzle). In that world, it would be possible to

redefine the units to ensure a constant price level in all periods, *without adding any uncertainty at all to firms' decisions*. If, however, firms' prices or quantities are uncertain, as in 99 percent of the world as we know it, then forcing a firm or household to announce a fixed real plan before the uncertainties are resolved generally results in a deterioration of a firm's ex ante profits or a household's ex ante utilities. Thus the more that firms are faced with market uncertainties that might lead them to change relative prices within the control period, the more the request-control system will pinch them. Nevertheless, as a stylized view of what a control system should do—as opposed to how it will be done—Baily's model looks ideal.

Christopher A. Sims: The “new contract theory” that Baily helped to develop has yielded valuable insights, particularly into the question of why some prices tend to be rigid even in the face of substantial quantity fluctuations. This theory is, however, very much a partial-equilibrium theory at its present stage of development, and in trying to use it, as in this paper, to generate macroeconomic conclusions, Baily has, in my view, put more weight on this fragile new apparatus than it can bear.

The theory described in the paper applies to a single employer and is static. There is no explanation of how prices and wages interact in an economy in which implicit and explicit contracts link many kinds of buyers and sellers. There is no explanation of how contract length is determined, or of just how the need to maintain a “reputation” allows implicit contracts to work without frequent default. These aspects of reality, taken as given in the new contract theory, are critical to macroeconomic applications.

Baily has presented a simulation model that seems to imply, for example, that mandating renegotiation of wage contracts at annual intervals would substantially reduce the unemployment fluctuations needed to control inflation. But because the simulation model does not generate contracts endogenously, one cannot measure the welfare losses that such a mandated form for contracts would evidently impose. Further, it should be apparent that every important part of the dynamics of the model would be affected by mandated changes in contract length. To the extent that sectoral labor-market conditions become less persistent as contracts shorten, the returns to workers from switching sectors decrease. To the extent that fluctuations in aggregate unemployment become less persistent, the response of the wage to aggregate demand might decrease. (That the

model allows no direct effect of *aggregate* unemployment on sectoral wage settlements is an evident defect.) To the extent that wages become less stable, the reaction of prices to wages would probably become more sluggish. This list could go on.

The conclusion that shorter contracts are likely to make inflation more sensitive to aggregate demand, at least in the short run, seems plausible to me. But if we were to try to get this model to do more than act as a ventriloquist's dummy, and instead, to predict realistically the size of the effect and the possible side effects, the number of modifications and sensitivity tests required would be prohibitive. The dummy would begin to behave like a tar baby.

The scheme of price controls Baily proposes appears to me to offer no real advantage over methods that have been used in the past. The ostensible advantage is that firms and workers can be allowed to set prices in nominal terms, and the control authority can then choose a single aggregate scaling parameter— α in Baily's notation. Such a mechanism apparently forestalls the tremendous load of microeconomic decisionmaking generally regarded as necessary to control aggregate price levels without distorting relative prices as allocational tools. However, crucial to this scheme is the condition that firms and workers will have no incentive to second-guess the aggregate scale parameter. For three reasons this condition does not hold in practice. First, if not all prices and wages are set synchronously, firms and workers will be concerned about their competitive position relative to others whose prices will remain fixed. Second, if "auction market" prices are left uncontrolled, firms will be concerned about their positions relative to them. (The passthrough of costs suggested by Baily does not solve this problem, since auction-market commodities may be substitutes for, as well as inputs into, contract-price goods.) Third, firms will second-guess α because of suspicions that aggregate-demand management will fail exactly to validate the controls. To these problems add the need to verify "value-added price" accounting. The conclusion must be that, in practice, the system would have to deal with requests for nominal-price changes on a case-by-case basis, just as control systems always have.

Finally, I think Baily makes too much of the effect on the new contract analysis of the introduction of flexible hours. His own analysis does make layoffs look much less likely in this case, but only because he considers

variable hours under a very special form of utility function in income and leisure. The form he chooses—linear in income and additively separable—implies that the dollar value of an hour of leisure to a worker can vary widely while the marginal utility of income is held constant. If utility functions concave in both income and leisure, and not additively separable, are admitted, one can find solutions in which neither hours nor the wage vary much with the state of the world but layoffs occur. The conditions on the utility function required to produce this result are a priori no more implausible to me than Baily's assumptions. Before throwing out what appears to me to be a basic insight of the new contract theory—that wage arrangements probably involve a natural form of insurance by employers of employees—I think the implications of the two sets of restrictions on utility functions ought to be put to a test.

This point is relevant to Baily's discussion of the importance of collusion in explaining layoffs, since a good part of the weight Baily attaches to collusive effects depends on his judgment that without collusion the limited observed variations in hours are hard to explain. Besides the fact that I don't share that judgment, I have some difficulty with Baily's argument as to why collusion should tend to make layoffs more likely. Unless group A workers and group B workers *prefer* layoffs to hours reductions, there is no reason a collusive contract could not provide for variation in hours. Baily claims that variations in hours give the employer incentives to default on the expected earnings differential between groups A and B. His reasoning appears to me to rest on an assumption that either the workweek or wages must be the same for all workers. I think that a contract that, for example, simply specified that all reductions in labor input would first affect group B workers, whether through layoffs or reductions in the workweek, would avoid the kind of default incentive that concerns Baily. The fact that less senior workers are commonly laid off in slack periods rather than put on short weeks must be explained by worker preferences or technology, not by the presence or absence of collusion.

Martin Neil Baily: Christopher Sims is certainly right in pointing out the many limitations of contract theory. I cannot explain just how implicit contracts work, or why explicit contracts take just the form they do. But I do believe that these contracts are an important phenomenon and that I can appraise some of their effects—even though I can't account for their

precise character. It is clear that firms do care about their reputations as employers, and that that concern influences their behavior.

I cannot understand the basis for Nordhaus' assumption that, when contracts are negotiated, the state of the labor market determines a rate of *deceleration* in annual wage increases, rather than an ongoing rate of wage increase. I don't believe that actual contracts embody the deceleration to which Nordhaus attributes compelling logic.

I can accept Sims' reservations about the application of partial-equilibrium microanalysis to macroeconomic problems. But I don't like any of the alternatives either. In particular, the simple extrapolation of aggregate relationships has often led us astray. One memorable example is the huge error made thirty years ago in predicting the postwar consumption function from the time-series data of the interwar period. Another is the extrapolation into the seventies of the previous stability of the macroeconomic relationship expressed by the Phillips curve. Clearly, neither aggregate nor microeconomic relationships give perfect answers; but used carefully, both can yield evidence on the puzzles that confront economists.

On the control scheme, I want to emphasize that I intended to offer a theoretical investigation of the conditions under which wage and price controls would function optimally—without creating distributional or allocational distortions. I did not present the model as a practical proposal for current implementation. But I do feel that some of the criticisms make the model sound more unrealistic than it is. For example, I believe that second-guessing the scaling parameter has been overemphasized by the discussants as a limitation on the relevance of the model. If any firm tries to second-guess α , it can suffer a market penalty by ending up with a relative price that is too high or too low. The virtue of the system lies in its enabling the market to continue to work. Nordhaus exaggerates the dislocation that would be caused by requiring fixed average markups from large corporations for three to six months. As they operate now, these firms reset prices only occasionally. The control system would surely not require Arrow-Debreu futures markets. On the contrary, its strength is its flexibility.

On one final point raised by Sims, I clearly could alter the labor-leisure utility function and account for layoffs without invoking collusion. But in that case I would really say, in one sense or another, that people get laid off because they like to watch TV or take vacations: that explanation doesn't seem very plausible to me.

Discussion

Robert Hall, Michael Wachter, and Charles Holt all agreed that contract theory was a very positive development. Hall, however, pointed to a number of loose ends in the current theory. One of these concerns is the way wages are set at contract time. Is it just the workers who happen to be employed in the firm at the time of the contract whose wishes are reflected in the bargaining? Or are laid-off workers represented in some fashion when the contract is negotiated? Second, he called for an answer to a recent critique of contract theory offered by Robert Barro. Barro argues that, if the purpose of the contract is to ensure *ex post* equality between the marginal product of labor and the marginal value of workers' time, a purely monetary disturbance should have no real effects, since it changes both of these factors by the same proportion. In that sense, Barro argues, contract theory does not explain why the short-run Phillips curve is not vertical.

Franco Modigliani observed that a monetarist, rational-expectations formulation would imply that the real effects from contracts could not extend beyond the horizon of the longest outstanding contract. Baily's results attribute an enduring effect to contracts because of the backward-looking price term in equation 8. Modigliani agreed strongly with Baily that past prices are a determinant of wages. He felt that people do in fact look backward, and that catching-up aspirations add to the momentum of the inflationary process. The man in the street looks to the past because he has no firm idea about the future. Agreeing with Modigliani, Arthur Okun distinguished between adaptive expectations and adaptive behavior. Because people recognize that they cannot forecast well, they adopt a practice of seeking compensation for what happened in the past; and that practice works out reasonably well because of the long-lasting attachments that mark a contract world.

Several participants expressed doubts that Baily's collusive model was necessary to explain either the prevalence of layoffs (rather than work-sharing or wage-cutting) in recession or the seniority-related nature of layoffs. Wachter suggested that the observed phenomena could be explained without invoking collusion by a life-cycle model with sorting and specific training. William Poole elaborated on the life-cycle aspects, arguing that the costs of being laid off were greater for more mature work-

ers, who have greater responsibilities than do younger people. Holt felt, however, that younger people might have greater burdens since they had to make heavier investments in durable goods and had maximum mortgage commitments. Poole responded that, if seniority-related layoffs were merely the result of collusion, new firms could find it profitable to offer employees a pledge that the hours of all workers would be reduced by the same proportion whenever employment had to be contracted. Finally, he suggested that commuting costs might be one element in the real world that argued against shortened workdays. Hall noted that, when a firm shut down for a full week, it was not making seniority-related layoffs. Yet that kind of work-sharing showed up as layoffs in the data rather than as a shorter workweek.

Agreeing with Sims' comments on the special properties of Baily's utility function involving leisure, Okun pointed out that, according to Baily's function, people would take time off exclusively in shorter workweeks and never in full-week vacations. He suggested that any utility function that accounted for vacations would also account for layoffs. Martin Feldstein agreed that utility functions that implied a disutility of a lumpy sort associated with working would be enough to account for some layoffs—apart from collusion, unemployment-insurance effects, or anything else. But Feldstein expressed his preference for an explanation of layoffs (rather than wage cuts) that relied on asymmetry of information about demand conditions between firms and employees. Workers would be skeptical if a firm informed them that it was cutting wages because demand had collapsed; but when the firm reduces employment and production, the workers know that it is telling the truth.

There was some discussion of what role the unemployment-insurance system might play in exacerbating layoffs. Okun pointed to the figures for 1929–32 cited in Baily's footnote 27 as a clear demonstration that the basic predominance of layoffs over shorter workweeks could not be attributed to either unemployment insurance, which was nonexistent at the time, or to unions, which then were extremely weak in manufacturing. Stephen Marston applied the argument to current conditions; he reported that only 8 percent of all firms are at their maximum unemployment tax rate, and 26 percent are at the minimum; thus, two-thirds of all firms face continuously increasing taxes if they lay off workers. Feldstein, however, cautioned that even for those firms the tax exemption of unemployment-insurance benefits provides an incentive for layoffs; and he suggested that

the marginal tax rate is about 30 percent for average workers. Feldstein also cited a study by Joseph Becker that found that in several states workers laid off by firms at their maximum unemployment tax rate accounted for half of all benefits.

While accepting Baily's description of his price-wage control scheme as a model of an ideal system, several participants wished to emphasize how far that ideal was from the real world. Wachter stressed that, because of the existence of nonsynchronized long-lasting contracts, the initiation of any controls system would inevitably disturb income shares and could not meet the requirement of distributional or allocative neutrality. Hall argued that the Baily plan involved greater administrative complexity than did the 1971–73 controls in one important respect; in the actual control system, only exceptional price increases had to be submitted, while Baily's scheme required all price actions to be ratified. Hall also expressed his reservations about any plan that enforced downward as well as upward price rigidity—an essential feature of Baily's system. Poole and Feldstein felt that firms would have to second-guess their rivals' decisions and that the plan would thus entail tremendous uncertainties for firms and induce complex price-setting strategies.

Martin Baily responded to several points raised in the general discussion. He rejected the distinction between monetary and real disturbances made in the Barro analysis that Hall had cited; he stated that contract theory deals with the responses of firms and workers to disturbances perceived by them as *real*. He also defended his disutility-of-work function, arguing that if the time period is viewed as a quarter rather than a week, the function is compatible with a desire for vacations. He acknowledged that short-term plant closings, such as Hall had mentioned, should then be interpreted as work-sharing (like hours reduction); but the rarity of widespread job rotation (one week off, two weeks on) remained a mystery unless collusion was present. Furthermore, he cautioned against putting too much weight on the value of leisure in a theory of recession layoffs.

In reply to Hall's doubts about restricting price reductions through the control system, Baily noted that prices could fall between request dates when costs of raw materials fall, and markups could be reduced by means of new, lower, requests. Allowing firms to preannounce their requests would eliminate some of the uncertainty that concerned Poole and Feldstein. Finally, Baily emphasized that the uncertainty caused by controls should be balanced against the uncertainty caused by recession.