Labor-Market Adjustments and the Persistence of Unemployment

By BRUCE C. GREENWALD AND JOSEPH E. STIGLITZ*

Persistent unemployment, like that plaguing Europe since the early 1980's, has been a persistent problem for economic theory. Competitive equilibrium theory *assumes* that all markets clear, including the labor market. All theories of unemployment thus must reflect significant departures from that paradigm. The last 20 years have generated a plethora of such theories. The challenge is to construct models that generate unemployment and are broadly consistent with a host of other labor and macroeconomic phenomena, including patterns of real wages and hours.

The traditional approach is to focus on a simple static equilibrium in which wages are kept above their market-clearing level for a variety of reasons: in the older versions of this story minimum wages, union power, and normative traditions; in its more recent incarnations, efficiency-wage considerations.[1] Within the United States, the older variants of these models have received decreasing credence, as union power has eroded, the real value of the minimum wage has declined, and empirical evidence has buttressed a broader set of theoretical arguments (based on imperfect competition within the labor market and efficiency-wage considerations) suggesting at most negligible effects from these government interventions. Efficiency-wage variants are more broadly consistent with observed labor-market behavior and help explain both natural rates of unemployment and certain cyclical phenomena (e.g., the use of layoffs rather than job-sharing) (see Carl Shapiro and Stiglitz, 1984; Lawrence Summers, 1990). Moreover, to the extent that unemployment compensation and protection against separations for cause have improved recently in Europe, these models can explain the secular rise in European unemployment (see Edmund Phelps, 1994).[2] On the other hand, as conventionally formulated, they leave many aspects of the labor market unexplained.[3]

More-recent theories analyze unemployment as the result of imbalances between flows into and out of the job market.[4] In these models, in steady state, on average, the flow supply into unemployment must balance the flow out. The equilibrium level of unemployment is thus determined through relations between these flows and the level of unemployment. Most noted within this genre are the labor-turnover theories, in which labor costs are affected by wages and the unemployment rate, and the job-search models, in which unemployment arises from the need for job matches, which by *assumption*, can only occur among the

*Greenwald: Department of Economics, Columbia University, New York, NY 10027, Stiglitz: Council of Economics Advisors and Department of Economics, Stanford University, Stanford, CA 94305. The views presented are those of the authors alone, and not of any institution with which they are or have been affiliated

[1] This older form of tradition goes back to Alfred Marshall. For surveys of the more recent literature, which seeks to endogenize the wage rigidities, see Stiglitz (1986).

[2] Note that these causes are very different from increased payroll taxes, which are often blamed for higher European unemployment since these would lead to lower wages, not higher wages and higher unemployment, as observed

[3] For example, the basic efficiency-wage model suggests greater cyclical wage sensitivity than is in fact observed (and has little to say about long-run changes in wages among demographic groups [e.g., college graduates] without corresponding changes in unemployment).

[4] For early examples of this literature, see Stiglitz (1974), and Dale Mortensen (1970).

unemployed.[5] The shortcomings of these search models are fourfold. First, to generate unemployment, they require that search when currently employed is far less efficient than job search while unemployed, which contradicts both much conventional wisdom and the fact that wage gains are typically far lower in work–unemployment–work transitions than in work–work transitions. Second, the search theory fails the critical empirical test presented by the transition from war to peace during 1945–1946 when search without intervening unemployment was unusually difficult (military service being in most cases incompatible with effective search), yet transitional unemployment was minimal (under 5 percent in 1946). Third, the search theories do not explain the failure of search intermediaries to alleviate the seemingly high social costs of the seemingly random interaction processes postulated by the models. Perhaps most importantly, they do not explain why wages do not adjust to alleviate the impact of shocks on either the flow demand or supply. Analyses of such adjustments are at the heart of understanding cyclical patterns of employment and unemployment.

I. Firm Adjustment Behavior in General

The model from which firm adjustments will be derived differs from conventional models of firm behavior in one critical respect. We will assume that firms are risk-averse and, therefore, maximize the expectation of a concave function of end-of-period firm value. Formally, firm decision-makers solve

$$\max E_t[u(W_t)] \tag{1}$$

where W_t is the terminal value of the firm (conditional on the decision-maker's information) and

$$W_t = W_{t-1} + \tilde{\pi}_t(\mathbf{x}_t, \tilde{\mathbf{z}}_t) \tag{2}$$

where W_{t-1} is inherited beginning-of-period value and $\tilde{\pi}_t$ is a random profit level for period t, which depends upon a vector of firm decision variables, $\mathbf{x}_t$, and a vector of environmental variables, $\tilde{\mathbf{z}}_t$, with

$$\tilde{\mathbf{z}}_t = \bar{\mathbf{z}}_t + \tilde{\boldsymbol{\varepsilon}}. \tag{3}$$

These latter variables are not known with certainty at the beginning of period t, when the levels of $\mathbf{x}_t$ are set,[6] but their expected values $\bar{\mathbf{z}}_t$ are assumed to be known.

The justifications for the assumption of risk aversion are firmly rooted in modern theories of imperfect information. For entrepreneurial firms, adverse selection in equity markets will prevent the full diversification of risks through equity sales and owner-managers will be left to maximize a function of the form of equation (1) (see Hayne Leland and David Pyle, 1977). Professional managers who may credibly threaten to reduce corporate value by a fraction, α, in any takeover contest will effectively own a fraction of firm value, and since this wealth is nontransferable, they too will maximize a function of the form of equation (1), with the α subsumed in the definition of the utility function. In a more conventional agency context, if principals are for practical reasons restricted to payment schemes that are linear in terminal firm value, the agent-managers' objective functions will again fit equation (1) (with the constant and scale terms of the payoff function subsumed in the utility function). Finally, models of net value maximization which include an appropriately defined managers' cost of bankruptcy lead to behavior entirely equivalent to that embodied in

[5] For instance, in labor-turnover models (e.g., Stiglitz, 1974), quit rates depend on wages and the unemployment rate. Given an unemployment rate, firms choose a wage to minimize total labor costs, and given those total labor costs, there is a particular demand for labor. The demand for labor plus the unemployed equals the labor supply. This basic framework can obviously be elaborated to take into account labor-market heterogeneity, vacancies, and so forth. See also Mortensen (1970).

[6] The variables $\mathbf{x}_t$ may include investments which add to stocks of a variety of capital variables without altering the implications of this model. For expositional convenience, however, this will not be made explicit in what follows.

equation (1) (see Greenwald and Stiglitz, 1993).[7]

The major consequences of this formulation for the adjustment behavior of firms are twofold. First, firm wealth effects, which are zero in the conventional theory of the firm (since conventional firms are assumed to be able to acquire capital freely in financial markets), play a critical role in the behavior of our risk-averse firms. To illustrate how significant these wealth effects may be, consider the simple case in which $\tilde{\pi}$ is characterized by constant returns to scale in the decision variables, $\mathbf{x}_t$, and u is characterized by constant relative risk aversion. Then the optimal levels of the decision variables will be linear in initial wealth. A 10-percent increase in initial firm wealth (equity) will imply a 10-percent increase in levels of investment and output.

Second, uncertainties concerning the impacts of different decision variables on $\tilde{\pi}_t$ will play a critical role in determining which variables are adjusted most significantly in the short run. If we assume further that uncertainties about the impact of decision variables increase with the distance from their current levels, the profit function can be written as

$$\tilde{\pi} = \pi\big[\mathbf{z}_{t-1}, \mathbf{x}_{t-1}; \tilde{\boldsymbol{\mu}}_t(\mathbf{x}_t - \mathbf{x}_{t-1}), \tilde{\boldsymbol{\eta}}_t(\tilde{\mathbf{z}}_t - \mathbf{z}_{t-1})\big].$$

The random variables $\tilde{\boldsymbol{\mu}}_t$ and $\tilde{\boldsymbol{\eta}}_t$ represent the *uncertainties* associated with both the decision variables $\mathbf{x}_t$ and the environment changes $\tilde{z}_t$.[8] If we linearize the π function and assume that $\tilde{\boldsymbol{\mu}}_t, \tilde{\boldsymbol{\eta}}_t$, and $\tilde{\boldsymbol{\varepsilon}}_t$ are independently and normally distributed, then efficient adjustments in $\mathbf{x}_t$, which can be defined as those which minimize the variance of $\tilde{\pi}$ subject to a constraint that the expected value of $\tilde{\pi}$ exceed a given level $\bar{E}$, will take the form

$$(4)\qquad \Delta\mathbf{x}_t^* \equiv \mathbf{x}_t^* - \mathbf{x}_{t-1}$$
$$= \tfrac{1}{2}\lambda\mathbf{v}_\mu^{-1}\bar{\boldsymbol{\mu}}_t - \mathbf{v}_\mu^{-1}\mathbf{c}'_{\mu\eta}(\bar{\mathbf{z}}_t - \mathbf{z}_{t-1})$$

where $\bar{\boldsymbol{\mu}}_t$ is the mean of $\tilde{\boldsymbol{\mu}}_t, \mathbf{v}_\mu$ is its covariance matrix, $\mathbf{c}'_{\mu\eta}$ is the transpose of the covariance matrix of $\tilde{\boldsymbol{\mu}}$ and $\tilde{\boldsymbol{\eta}}$, and λ is a multiplier associated with the minimum expected profit constraint.

The form of these optimal adjustments is actually far less complicated than it may seem. The first term on the right-hand side of equation (4) represents the *active* or profit-enhancing changes undertaken by the firm. In the simple case where the instrument uncertainties are independent and hence $\mathbf{v}_\mu$ is diagonal, the value of this term for the ith instrument is

$$\Delta x_{it}^* = \tfrac{1}{2}\lambda\frac{\bar{\mu}_{it}}{\sigma_{it}^2}$$

where σ_{it}^2 is the variance of the instrument uncertainty associated with the ith decision variable. This is just the Sharpe ratio of expected return to variance resulting from the use of that variable. Other things being equal, in the full portfolio of decision-variable changes, adjustments will be concentrated in variables with certain effects (low σ_i^2) and large expected returns. The constant λ, which multiplies all active changes, traces out points on the frontier of mean–variance possibilities. In general, the less risk-averse or, with declining risk aversion, the greater the initial wealth of the firm in question, the greater λ will be, and the more aggressively the firm will pursue enhanced profit opportunities.

The second term in equation (4) is a *passive* adjustment to environmental changes designed to minimize the net impact of those changes on profit uncertainty. The elements of the vector $\mathbf{v}_\mu^{-1}(\mathbf{c}_{\mu\eta})'_i$, which multiply the changes in $(\bar{\mathbf{z}}_{it} - \mathbf{z}_{it-1})$ to yield the passive changes in the i decision variables, are effectively the coefficients of a regression of

[7] Note that in agency models nonlinear payment schedules lead to maximization problems of the form of (1), except when the principal is attempting strongly to encourage risk-takings.

[8] The first of these is called instrument uncertainty and is similar to that identified in the context of macroeconomic policy formulation by William Brainard (1967).

the jth environmental uncertainty on the instrument uncertainties of the i decision variables. Taken together they minimize the net impact of the environmental variable on the overall profit function. The obvious example of this is a situation where z_j is the economy-wide price level and x_i is the own price for a particular product. If there is no money illusion, then the effects of the price level and own price will be equal and opposite, although neither is known with certainty. Then the regression of the price-level effect on the own-price effect will be -1, and the appropriate passive adjustment to an expected level of inflation will be full incorporation into the own-price variable, before considering active real-price adjustments. This constitutes the broad theory of firm-level adjustment that we apply to the labor market.[9]

II. Firm Adjustment Behavior in the Labor Market

In applying the basic structure of the previous section to labor-market adjustments, we will use a very simple basic model. We assume that labor is the only input to production and that there are constant returns to scale. Total labor input will consist of the product of a number of workers, L_t, the number of hours worked per worker (assumed for the moment to be the same for all workers), h_t, and an average level of worker effort, e_t, which depends for efficiency-wage reasons on hours worked, real wages (w), and the extent of new hiring or dismissals. In addition, we will assume that there are costs associated with either increasing or decreasing employment which will be designated $m(\Delta L_t)$, where ΔL_t is the change in employment. These embody hiring and training costs when $\Delta L_t > 0$ and layoff cost when $\Delta L_t < 0$. Finally, we will assume that the firm's terminal labor force has a capital value $\tilde{k}_t$ per unit which is a random variable at the time hiring decisions are made. Under these conditions,

$$W_t = W_{t-1} + \left[\tilde{e}_t(w_t, h_t, L_t) - w_t\right] h_t L_t - m(L_t - L_{t-1}) + \tilde{k}_t(L_t - L_{t-1}) + \left(\tilde{k}_t - k_{t-1}\right) L_{t-1}$$

is the firm wealth variable that management will be assumed to maximize, where W_{t-1} includes the value of the inherited labor force L_{t-1} at a known value per unit k_{t-1}.

There are two further simplifying assumptions that we will make. First, we will ignore the passive adjustment component of the optimal policies identified in the previous section. In practice this means (i) assuming that firms effectively set real wages (i.e., the comparative wage-level shifts are fully embodied in the firm's wage policy); (ii) assuming that firms properly anticipate random turnover which we will assume is zero;[10] and (iii) not attempting to model explicitly firm responses to changes in nonwage external variables like the unemployment rate. The second simplifying assumption is that instrument uncertainty affects only the expected effort level of workers. Thus, after linearizing,[11]

$$(5)\quad \widetilde{W}_t = W_{t-1} + \left[\tilde{\mu}_{ew}(w_t - w_{t-1}) + \tilde{\mu}_{eh}(h_t - h_{t-1}) + \tilde{\mu}_{eL}(L_t - L_{t-1}) + e_{t-1} - w_t\right] h_t L_t - m(L_t - L_{t-1}) + \tilde{k}_t L_t - k_{t-1} L_{t-1}$$

[9] This brief summary elides two extensions that we allow in the labor-market case. In particular, we allow the possibility that for some decision variables uncertainty increases with the level of the decision variable itself, not just the deviation from past values (as will become apparent, this applies chiefly to capital variables), and the wealth variable includes not only "cash," but other state variables

[10] The effect of this assumption is to rule out an intermediate hiring decision between positive hires and layoffs, represented in practice by freezing hiring and allowing L_t to fall through natural attrition. With the present assumption, this corresponds to a zero hiring level

[11] To see this, note that

$$\tilde{e}(w, h, L) \approx \tilde{e}(w_{t-1}, h_{t-1}, L_{t-1}) + \tilde{\mu}_{ew}(w_t - w_{t-1}) + \tilde{\mu}_{eh}(h_t - h_{t-1}) + \tilde{\mu}_{eL}(L_t - L_{t-1}).$$

and, for completeness,

$$m(L_t - L_{t-1}) = \begin{cases} d_{1t}(L_t - L_{t-1}) & \text{if } L_t > L_{t-1} \\ 0 & \text{if } L_t = L_{t-1} \\ -d_{2t}(L_t - L_{t-1}) & \text{if } L_t < L_{t-1} \end{cases}$$

which is always positive.

Efficient labor-market adjustment for a risk-averse firm (assuming that $\tilde{\mu}_{ew}, \tilde{\mu}_{eh}, \tilde{\mu}_{eL}$ [the random variables defining the impact of wages, hours and force levels on effort] and $\tilde{k}$ are independently and normally distributed) will be ones which solve

$$\min \operatorname{Var}(\widetilde{W}_t)$$

subject to that constraint that $E(\widetilde{W}_t) \geq \overline{E}$ for appropriate levels of $\overline{E}$. For hours and wage changes, these take the form

$$(6) \quad \Delta w_t^* = \tfrac{1}{2}\lambda(\overline{\mu}_{ew} - 1)h_t L_t / \sigma_{ew}^2 h_t^2 L_t^2$$

and

$$(7) \quad \Delta h_t^* = \tfrac{1}{2}\lambda\left[\overline{\mu}_{eh} + \left(\frac{\bar{e} - w_t}{h_t}\right)\right] h_t L_t / \sigma_{eh}^2 h_t^2 L_t^2$$

which are particular versions of the general formulas of the previous section with σ_{ew}^2 and σ_{eh}^2 being the variances of the effects of wages and hours, respectively, on effort and $\overline{\mu}_{ew}$ and $\overline{\mu}_{eh}$ being the expected effects of wages and hours on effort.[12] Thus,

$$(\overline{\mu}_{ew} - 1)h_t L_t$$

is the expected profitability increase from an increase in wages, and

$$[\overline{\mu}_{eh} + (\bar{e} - w_t)/h_t]h_t L_t$$

is the expected profitability gain from an increase in hours worked.

Here the critical issue is the relative impact of hours and wages on worker effort and efficiency. If changes in hours (paid for at the existing wage) are widely accepted and, therefore, have little impact on effort, then the instrument uncertainty associated with hours adjustments will be small (since $\bar{e}$, w_t, and h_t are likely to be observed with little or no error). If σ_{eh}^2 is, therefore, much smaller than σ_{ew}^2, short-term adjustment will be heavily weighted to changes in hours rather than changes in wages. The greater instrument uncertainty associated with wage changes will lead to smaller initial adjustments which continue for longer periods (since the expected gains from wage adjustments will not be as rapidly driven to zero).

Optimal labor-force adjustments, L_t, are more intricate. The derivative of the Lagrangian of the constrained maximization problem which defines efficient labor-force adjustments is, ignoring squared and higher-order terms,

$$\frac{d}{d(\Delta L_t)} = 2\sigma_{el}^2 h_t^2 L_t^2 \Delta L_t + 2\sigma_k^2 L_t - \lambda\left[\overline{\mu}_{el} h_t L_t + (\bar{e} - w_t)h_t - m' + \overline{k}\right].$$

This must be less than or equal to zero. Thus,

$$(8) \quad \Delta L_t^* \leq \frac{1}{2\sigma_{el}^2 h_t^2 L_t^2}\left\{\lambda\left[\overline{\mu}_{eL} h_t L_t + (\bar{e} - w_t)h_t - m' + \overline{k}\right] - 2\sigma_k^2 L_t\right\}$$

where $m' = d_{1t}$ if $\Delta L_t^* > 0$ and $m = -d_{2t}$ if $\Delta L_t^* < 0$.

Equation (8) can easily be interpreted. Let $a = [\overline{\mu}_{eL} h_t L_t + (\bar{e} - w_t)h_t + \overline{k}]$ be the value of increased mean terminal value (profits plus mean value of labor stock) associated with higher employment levels; similarly, let $b = 2\sigma_k^2 L_t / \lambda$ be the "risk cost" associated with having a larger labor force. Equation (8) says to hire additional laborers if

$$a - b > d_{1t}$$

(the gains from hiring are greater than the

[12]These expressions assume that changes in h_t and w_t are relatively small. We ignore second-order and higher terms in Δh_t and Δw_t, also assuming that the impact of these changes on $\bar{e}$ is small.

cost of adjustment); fire workers if

$$a - b < -d_{2t}$$

(the losses exceed the cost of firing); and neither hire nor fire if

$$d_{1t} > a - b > -d_{2t}.$$

The firm is in one of three regimes: hiring, firing, or doing neither. In the intermediate regime, firms rely primarily on hours adjustments accompanied by slower, steadier adjustments in wages (assuming there is greater risk associated with these wage adjustments). In the other two regimes, the brunt of adjustment is borne by changes in the labor force, accompanied by slower, steady wage and hours adjustments. As b decreases, the firm switches from firing, to neither firing nor hiring, to hiring.

The critical observation for our purposes is that *decreases in firm net worth increase the firm's risk aversion, reduce λ and thus increase b; and increases in economic uncertainty increase b.* Thus, as firms' net worth decreases, their tolerance for risk decreases, and they may switch from hiring workers, to neither hiring nor firing (allowing hours and wages to take the brunt of adjustment), to engaging in layoffs. In the context of economic downturns, systematic patterns of decreased firm net worth are typically associated with increased perceptions of risk and decreased expected returns at given wages and hours, reinforcing the changes in regimes. As more firms switch from a hiring regime to the no-hiring/no-layoff regime, or from that regime to the layoff regime, unemployment obviously increases, since the flow out of employment will increase, and the flow into employment will decrease.

Over the course of a business cycle, firms tend to move from the hiring regime (at the peak) to the intermediate regime to the layoff regime (in the trough) and back again. The sequence of observed adjustments should be the sequence of hiring reductions, followed by hours reduction, and only after an interval, layoffs; then, as the economy reverses course, hours will be extended, followed after an interval by renewed hiring. Wages, being the labor-input control variable with greatest instrument uncertainty, should change slowly and persistently over the course of the business cycle. This sequence appears to correspond closely to what is observed in practice.[13]

III. Unemployment Levels

The implications of the patterns of adjustment outlined above for persistent unemployment levels, such as those observed in Europe, depend upon the extent to which labor-market rules, on the one hand, expand the range over which firms adjust primarily by hiring workers and, on the other hand, shrink the range over which layoffs occur. Merely making layoffs prohibitively expensive will not increase employment and is, indeed, likely to be counterproductive. At the lower end of the scale of firm prosperity, firms that would like to lay workers off, but for legal or institutional reasons cannot do so, will ultimately go bankrupt, and layoffs will accompany the final act of dissolution. At the upper end of the firm prosperity scale, both the expected benefits to higher labor-force levels ($\bar{k}$) and, perhaps more importantly, the risks associated with inflexibly higher labor-force levels (σ_k^2) will deter hiring. On balance, therefore, after a short period of reduced layoffs and reduced hiring, layoffs will increase with no corresponding rise in hiring. Ultimately, of course, the two must come into balance if unemployment is not to increase without limit. However, the net result is likely to be that this balance will occur at a persistently higher level of unemployment. And where efficiency-wage considerations inhibit off-setting wage reduc-

[13]A second implication of the model that appears to correspond closely to observed practice is the concentration of adjustments among well-defined work groups. If the variables affecting effort put out by a well-defined group of workers are those which apply only to that group, then the uncertainties associated with labor-market adjustments can be minimized by applying the layoff wage and hours changes only to well-defined groups (e.g., by seniority, job classification, or location) or by restricting changes to well-defined time periods.

tions (or spread them over a long period of time), the rising unemployment occasioned by the imbalance between hiring and layoffs may be sustained for a considerable period of time.

REFERENCES

Brainard, William. "Uncertainty and the Effectiveness of Policy." *American Economic Review*, May 1967 (*Papers and Proceedings*), *57*(2), pp. 411–25.

Greenwald, Bruce and Stiglitz, Joseph. "Financial Market Imperfections and Business Cycles." *Quarterly Journal of Economics*, February 1993, *108*(1), pp. 77–114.

Leland, Hayne and Pyle, David. "Informational Asymmetries, Financial Structure and Financial Intermediation." *Journal of Finance*, May 1977, *32*(2), pp. 371–87.

Mortensen, Dale. "A Theory of Wage and Employment Dynamics," in Edmund Phelps et al., *Microeconomic foundations of employment and inflation theory*. New York: Norton, 1970, pp. 167–211.

Phelps, Edmund S. *Structural slumps: The modern equilibrium theory of unemployment, interest, and assets*. Cambridge, MA: Harvard University Press, 1994.

Shapiro, Carl and Stiglitz, Joseph E. "Equilibrium Unemployment as a Worker Discipline Device." *American Economic Review*, June 1984, *74*(3), pp. 435–44.

Stiglitz, Joseph. "Alternate Theories of Wage Determination and Unemployment in LDC's." *Quarterly Journal of Economics*, May 1974, *88*(2), pp. 194–227.

______. "Theories of Wage Rigidity," in J. L. Butkiewicz, Kenneth J. Koford, and Jeffrey B. Miller, eds., *Keynes' economic legacy: Contemporary economic theories*. New York: Praeger, 1986, pp. 153–206.

Summers, Lawrence. *Understanding unemployment*. Cambridge, MA: MIT Press, 1990.