

THE EFFICIENCY WAGE DEBATE

1. Introduction

Post-Walrasian macroeconomists criticize traditional Keynesian (IS-LM) macro theory for assuming away the micro dimensions of the economy via excessive aggregation, and they criticize new classical macro theory for assuming away the macro dimensions of the economy by resorting to the device of the single “representative” consumer and the single “representative” firm. They are especially critical of classical macroeconomists, however, for denying the very existence of coordination problems in decentralized market economies. In particular, they criticize the classical assumptions that markets are complete, that prices are fully flexible, and that all agents are price takers, but that prices nevertheless continually adjust to keep markets continually cleared.

One strand of the post-Walrasian literature, generally referred to as New Keynesian economics, claims to have developed a new theory of macroeconomics that relies on micro-based models of imperfect information. *Asymmetric information* in a market for goods, services, or assets refers to differences (“asymmetries”) between the information available to buyers and the information available to sellers. New Keynesian argue that asymmetric information is endemic in markets due to costs of information acquisition. This asymmetry leads to problems of adverse selection¹ and moral hazard² that are the root cause of many

¹*Adverse selection* is a pooling quality problem that arises for buyers when they have difficulty assessing the quality of items in advance of purchase. Roughly speaking, buyers lacking information about true quality are only willing to pay a price equal to the value of an average quality good; but this then induces sellers of high quality goods to withhold their goods from the market and sellers of low quality goods to bring their goods to the market. The result can be a “lemons problem” – a steadily decreasing quality level for the pool of goods available for purchase in a market.

²*Moral hazard* is said to exist in a market if, after the signing of a purchase agreement between a buyer and seller: (i) one of these agents changes his or her behavior in such a way that the probabilities used by the other agent to determine the terms of the purchase agreement are no longer accurate assessments of risk; and (ii) the latter agent is only imperfectly able to observe this change in behavior. For example, a moral hazard problem arises if, after a lender purchases a loan contract from a borrower, the borrower uses the

of the coordination difficulties perceived in labor and credit markets.

In particular, as an important part of their story, New Keynesians advocate “efficiency wage” theories of the labor market as developed by Solow (1979) and Akerloff and Yellen (1986), among others. The following subsections provide an introduction to various types of efficiency wage theories. As will be seen, these theories commonly assume that worker productivity depends positively on the real wage rate. Moreover, employers are assumed to have incomplete information about the quality of their employees, hence information is asymmetric. Finally, prices and wages are assumed to be set by agents in the economy rather than by some unexplained adjustment process. Consequently, imperfect rather than perfect competition is assumed, in the sense that agents use price and wage policies in an attempt to attract customers for their goods and services as well as providers of needed inputs for their production processes. Additional assumptions special to each theory then distinguish the various theories from each other.

2. Overview of the Efficiency Wage Theory

Why do wage differentials persist for apparently similar employees in apparently similar jobs? Why are unemployed workers unable to bid down the wages of comparably employed workers and thus gain jobs? Do these “sticky wages” correspond to rational economic behavior? Efficiency wage theories suggest that the answer to these questions may lie in the negative incentive effects of low wages on worker efficiency.

Basic Efficiency Wage Hypothesis: Worker efficiency depends *positively* on the real wage rate, where worker efficiency is defined as the intensity of effort exerted per hour of work.

The key implication is then as follows. If worker efficiency depends positively on the real wage rate, firms may find it profitable (“rational”) to pay real wages in excess of the real borrowed funds for investments that are riskier than reported to the lender and the lender is not able to monitor this behavior. The contractual interest rate may then be too low to cover the actual risks faced by the lender.

wage rates that clear the labor market; for the loss in efficiency of all workers currently on the job which results from a cut in real wages might more than offset the decrease in real wage costs and/or the incremental increase in output due to the hiring of more workers at the lower real wage. In this case there will be persistent involuntary unemployment.

3. An Illustrative Efficiency Wage Example

The example developed in this section is taken from Solow (1979). Consider a profit-maximizing firm that behaves as a perfect competitor in its output market (i.e., it takes its output price as given) but behaves as an imperfect competitor in the labor market (i.e., it sets its own real wage). The firm sets the same real wage for each of its employees. The short-run production function of the firm takes the form

$$q = aF(eL) , \tag{1}$$

where e is the efficiency level of each employee, L is labor measured in homogeneous person-hours, a is a productivity shift factor, q is output, and $F(\cdot)$ satisfies $F' > 0$ and $F'' < 0$.

NOTE: In old-style Marxian terminology, L is *labor power* measured in time, and represents the capacity to labor. On the other hand, *labor* takes into account the quality of the labor, and thus is analogous to eL in Solow's model. Marxians have long stressed that the ability to extract labor from labor power depends on the "social relations of production," which includes a consideration of "fair pay" as well as other types of work norms.

Output is taken to be the numeraire good with a price $P = 1$, taken as given by the firm. Let w denote the real wage, i.e., the wage measured in units of q :

$$w = [\text{Units of } q] / [\text{per Person-Hour}] . \tag{2}$$

Suppose the efficiency level e depends on the real wage w , as follows:

$$e = e(w) \text{ with } e' > 0 \text{ and } e(0) = 0 . \tag{3}$$

Suppose also that the elasticity of e with respect to w , i.e., $e'(w)w/e(w)$, is a declining function of w .

Suppose the firm believes it can hire all the labor it wants at the wage it chooses to offer. Then its profit-maximizing problem takes the form

$$\max_{w,L} [aF(e(w)L) - wL] \quad (4)$$

with respect to w and L , subject to $w, L \geq 0$. The first-order necessary conditions for a positive vector (w, L) to solve (4) take the form

$$aF'(e(w)L)e'(w)L - L = 0 ; \quad (5)$$

$$ae(w)F'(e(w)L) - w = 0 . \quad (6)$$

Assuming $L \neq 0$, one can divide through (5) by L and use (6) to substitute out for $aF'(e(w)L) = w/e(w)$. The result is

$$we'(w)/e(w) = 1 ; \quad (7)$$

$$MPL \equiv ae(w)F'(e(w)L) = w . \quad (8)$$

Let $(w^*, L^*) > 0$ denote a solution to these first-order conditions. By (7), the optimal wage w^* satisfies the “Solow condition” that the elasticity of efficiency with respect to the wage is unity. The wage w^* is known as the **efficiency wage** since it is the wage which minimizes wage costs per efficiency unit of labor, i.e., $[wL/e(w)L]$, or simply $[w/e(w)]$. The first order necessary condition for a wage w to minimize this expression, given $e(w) \neq 0$, is given by (7). By (8), each firm demands labor up to the point where the marginal product of labor just equals w^* .

Technical Note: Using (5) and the assumption that $F'' < 0$, it can be shown the labor demand curve (8), graphed in the $L - w$ plane, is everywhere downward sloping. That is,

$dw/dL < 0$ everywhere along this labor demand curve. More precisely, letting net profits associated with a choice of (L, w) be denoted by $G(L, w) = aF(e(w)L) - wL$, it can be shown that (5), (7), and (8) imply the existence of an *inverse* labor demand function $w = w(L)$ satisfying

$$w'(L) = -G_{11}(L, w(L))/G_{12}(L, w(L)) = -w/L < 0, \quad (9)$$

where

$$G_{11}(L, w) = ae(w)F''(e(w)L)e(w) < 0; \quad (10)$$

$$G_{12}(L, w) = ae(w)F''(e(w)L)e'(w)L < 0. \quad (11)$$

Moreover, it follows from (9) that the wage elasticity $-w'(L)L/w$ equals 1 everywhere along this inverse labor demand curve.

What about the sign of profits at the optimal point $(w^*, L^*) > 0$? Let $z^* \equiv e(w^*)L^*$, and note that $z^* > 0$ by assumption. Using the first order condition (8), and the assumed strict concavity of $F(\cdot)$, one then obtains

$$\begin{aligned} aF(z^*) - w^*L^* &= aF(z^*) - [ae(w^*)F'(e(w^*)L^*)]L^* \\ &= az^*[F(z^*)/z^* - F'(z^*)] \\ &> 0. \end{aligned} \quad (12)$$

Thus, profits are strictly positive.

As long as the aggregate demand for labor falls short of aggregate labor supply and w^* exceeds labor's reservation wage, the firm will be unconstrained by labor market conditions in pursuing its optimal policy (w^*, L^*) . Consequently, equilibrium will entail persistent involuntary unemployment. Unemployed employees will strictly prefer to work at w^* rather than be unemployed, but firms will have no incentive to hire them at that wage or to lower the wage. By construction, any change away from the profit-maximizing solution (w^*, L^*)

will decrease the firm's profits. In particular, any cut in w^* lowers the productivity of all employees on the job, and this lower productivity outweighs any benefit in terms of lower wage costs.

What about real wage rigidity?

Real wage rigidity at the firm or industry level – that is, the invariance of the optimal real wage w^* to the scale of production – is a straightforward implication of this model. Note that w^* is completely determined by (7), hence w^* only depends on the form of the efficiency function $e(w)$. Note, in particular, that w^* is entirely independent of the level of the productivity shift factor a . Rather, recalling that $F'' < 0$, it follows from (8) that the firm's response to a decrease in a is to lay off employees so that the MPL is still equated to w^* .

Solow (1979) shows that the real wage rigidity arises for a production function of the form $q = f(L, w)$ if and only if the real wage enters the production function in a “labor-augmenting” way — that is, if and only if the production function can be represented in the form $q = F(e(w)L)$ where F and e are any strictly increasing functions. A general production function of the form $q = f(L, w)$ need not generate real wage rigidity.

What about the unitary wage elasticity of efficiency?

With a different type of production relation, the elasticity of efficiency with respect to the real wage need no longer be unitary in the optimal solution (w, L) .

Akerloff and Yellen (1986, p. 15) give a simple example of this. Suppose that a *single* employee controls the water flow from a large privately owned dam. The employee has an inelastic supply of labor given by $L = 1$. The output q from the dam is proportional to the effort $e(w)$ exerted by this single employee:

$$q = ae(w) . \tag{13}$$

where $e'(w) > 0$. If the employee shirks (reduces e), the resulting loss in productivity cannot

be even partially compensated for by the increase in L from other employees.

Suppose the real value of a unit of output is 1. Then the dam's profits for each given wage rate w are given by

$$ae(w) - w . \tag{14}$$

The first order condition for a wage w^* to maximize profits is then

$$ae'(w^*) = 1 . \tag{15}$$

Assuming profits are positive in equilibrium, so that $ae(w^*) > w^*$, relation (15) yields

$$1 = \frac{ae(w^*)e'(w^*)}{ae(w^*)e'(w^*)} > \frac{w^*e'(w^*)}{ae(w^*)e'(w^*)} = w^*e'(w^*)/e(w^*) , \tag{16}$$

which implies that the elasticity of efficiency with respect to the real wage is less than one at the optimal solution point w^* .

Note how this model differs from Solow's model. In the latter model, the profit function F was a nonlinear function of L , and maximization with respect to L gave a finite solution. For the model at hand, the profit function is linear in L , hence the maximization of this profit function with respect to L for any given w would give an infinite solution for L unless $ae(w) - w = 0$, i.e., unless profits for that w happened to be zero.

4. Alternative Efficiency Wage Models

A. Segmented Labor Market Models

The *dual labor market hypothesis* states that the labor market can be roughly segmented into a "primary sector" that offers jobs characterized by high wages and internal labor markets (promotion ladders) and a "secondary sector" that offers low-paying menial jobs with little room for advancement. The efficiency wage approach can be used to provide a rationale for this segmentation.

Suppose that employees in one group of firms have great flexibility with regard to choosing the level of effort they can exert on the job, and increased work effort results in significant

increases in productivity (e.g. blue collar automobile factory employees organized in teams). On the other hand, employees in a second group of firms perform routinized tasks for which increases in effort would yield little in terms of increased productivity (e.g., blue collar automobile factory employees organized in assembly lines).

Then the efficiency wage w^* will differ across these firms, and a distribution of wages for employees with identical basic capabilities can arise in equilibrium. High wages (relative to the market clearing wage) and job rationing can then arise in the “primary” sector where efficiency-wage considerations are important, while wages close to the competitive wage may prevail in the “secondary” sector where they are less important.

Although every unemployed worker might be able to get a job in the secondary sector if he chooses to, and in this sense is “voluntarily” unemployed, it is also true that unemployed workers would be more than willing to work in primary sector jobs at the prevailing higher wage rates. In this sense they are “involuntarily” unemployed.

B. Moral Hazard/Shirking Models

If employers are unable to observe the efficiency-wage relation $e(w)$, moral hazard problems can arise in the labor market as employers seek with only partial success to monitor the work of their employees in relation to their wages. An employer may pay an employee a particular real wage in anticipation that the employee will produce a certain amount of output per hour, and he may announce that any employee caught shirking (i.e., exerting a level of effort below what the employer anticipates) will be fired. However, if there is some chance that shirking will not be observed by the employer, an employee may still choose to engage in some degree of shirking.

Employers may then have an incentive to pay high real wages to their employees in order to reduce shirking, since this raises the cost to the employee of losing his job. But the degree to which employers are willing to pay these high wages clearly depends on the cost

to employees of being fired in the absence of any wage bonus.

This brings out the need for more explicit microfoundations to explain the efficiency-wage relation. In particular, factors other than just the wage rate may affect an employee's efficiency level. For example:

- The expected income from unemployment may affect the efficiency-wage relation. For any given level of efficiency \bar{e} , a higher unemployment insurance benefit may raise the real wage required to achieve \bar{e} , and hence reduce employment, since the cost to the employee of being fired for shirking (low efficiency) is reduced.
- A higher unemployment rate in the economy, and hence a longer expected duration of unemployment for any employee fired for shirking, may reduce the wage needed to achieve \bar{e} . In this sense, unemployment acts as an employee disciplinary device (shades of Marx's "army of the unemployed").
- The level of wages offered by *other* employers affects the prospects of a fired employee, and hence may affect the efficiency-wage relation he maintains in his current place of employment. Thus, employers have to be concerned with their position in the wage hierarchy within their industry.

C. Adverse Selection Models:

Whenever buyers in a market can only observe the average quality of the goods being offered for sale, the market price level will tend to be set at a level appropriate for a good of average quality. Sellers of high quality goods will then tend to withdraw their goods from the market because they are not being rewarded for high quality, whereas sellers of low quality goods will tend to bring their goods to the market. The average quality of the goods in the market (and hence the market price level) will thus tend to drop over time. As noted in footnote 1, this process is referred to as *adverse selection*.

Adverse selection provides yet another reason for the existence of a relationship between productivity and real wages. Given heterogeneous workers, the particular pool of workers attracted to a firm may depend on the real wage offered by the firm, with higher real wages resulting in a pool of workers of higher average quality.

Suppose, for example, that performance on the job depends on “ability” b , and workers have heterogeneous levels of ability. Suppose also that the reservation real wage \bar{w} (i.e., the minimum real wage at which a worker is willing to work) is an increasing function of ability b . Then firms offering higher real wages will attract more able workers into their work pool along with less able workers.

If firms cannot observe applicant quality and lack devices to induce workers to reveal their true abilities—or if institutional or legal constraints prevent firms from hiring on the basis of ability—random hiring from the applicant pool must be done. A higher real wage increases the expected ability of a worker hired randomly from the applicant pool. A real wage above the market clearing level may minimize real wage costs per efficiency unit of labor under these circumstances.

In this case, $e(w)$ represents a relation between the real wage and the average productivity level of a firm’s pool of workers. It is not that each individual worker increases his effort in response to a higher real wage. Rather, the pool of workers itself changes in response to a higher real wage.

D. Turnover Models

Workers are less likely to quit a job the higher the relative real wage paid by their current firm and the worse the prospects in the external labor market (e.g., the higher the aggregate unemployment rate). If firms must bear part of the costs of turnover (e.g., because they have invested training or hiring costs in the worker which have not been recovered), and if quit rates are a decreasing function of real wages paid, firms have an incentive to pay high

real wages to reduce costly labor turnover—i.e., to pursue a high real wage/low turnover policy. Interestingly, in some cases it may also be profitable for the firm to pursue a low real wage/high turnover policy.

E. Union Threat/Insider-Outsider Models

The union threat model predicts that real wage premiums should arise where the costs of union organization are low for workers and where the potential gains from unionization are high. Insider/outsider models also generate wage premiums and job rationing based on the bargaining power of “insiders” (incumbant workers). See, for example, Lindbeck and Snower (1988).

5. Theoretical Critiques of Efficiency Wage Models

As intuitively appealing as these more general efficiency wage models may be, they suffer from a basic difficulty – only simple wage schemes are considered, with dismissal as the only penalty for shirking. In many cases, however, Pareto-superior forms of employment contracts exist which could eliminate the moral hazard and adverse selection problems.

One possibility is that workers are required to post **performance bonds** at the time of hiring as insurance to the firm that they will not shirk. If shirking is detected, the bond is forfeited. Alternatively, the workers and employer might agree in advance that workers caught shirking will have to pay a fine. The threat of having to forfeit a bond or pay a fine substitutes for the threat of being fired.

For example, a basic objection to the adverse selection model is that firms would eventually learn the ability levels of their workers, and so some kind of performance bond arrangement would eliminate the adverse selection problem without need for an efficiency wage.

Another possibility is to require workers to pay an **entrance fee** at the time of hiring. If firms use high wages to encourage workers not to shirk, these workers should be willing to pay the entrance fee to gain employment at the firm.

Another possibility is to have **seniority wages**. Workers can be paid a wage less than their marginal productivity when they are first hired, or be charged some type of fee, with a promise that their earnings will later exceed their marginal productivity if they are retained by the firm. This provides an incentive for workers not to shirk, because a worker fired for shirking forfeits the increased earnings promised to more senior workers.

6. Empirical Support for the Efficiency Wage Hypothesis

A standard competitive labor market model implies that persistent industry wage premiums require industry-related differences in labor quality or nonwage dimensions of the work such as location or risk—i.e., in the long run, wage differentials for similar types of work must be *compensating* differentials that essentially equalize the workers who undertake the work.

Empirical area wage studies invariably find large and persistent wage differentials across firms for a given job classification within a given geographical location. Substantial wage differentials remain even after controlling for union status and observed worker characteristics. Are these differentials compensatory or not?

Efficiency wage theories predict that there should be important wage differentials not explained by compensating differentials or by shifts in labor demand across sectors. As noted above, reasons given for these non-compensatory wage differentials include monitoring costs, turnover costs, high cost (risk) to firm for worker shirking, union threat, etc. To date, however, it has proved difficult to reconcile the observed wage differentials with any particular variant of the efficiency wage argument.

The main econometric issue is that the error term in the standard wage equations used to test for the presence of efficiency wage premiums (non-compensatory wage differentials) may contain unobserved ability and job-attribute components which are correlated with an individual's reported productivity. In this case, the resulting estimates for non-compensatory

wage differentials are biased upwards.

Indeed, skeptical researchers such as Murphy and Topel (1990) interpret the evidence as saying there are enormous unobserved ability differences among individuals and enormous unobserved job-attribute differences across industries. In other words, the competitive labor market is alive and well; its just that we haven't yet been able to identify the underlying compensatory reasons for the wage differentials.

7. Empirical Evidence on Contracts

In actuality, instances of workers posting bonds or paying entrance fees are rare. More common are pensions and other deferred compensation schemes, and internal promotion ladders. The evidence for seniority wages is not clear-cut.

One reason for this may be that workers, especially young workers, do not have enough savings or credit to be able to post a performance bond or pay an employment fee. If the probability of detecting shirking is low, the size of the required bond or fee can be substantial. Indeed, the empirical observation that firms do devote substantial resources to monitoring their workers suggests that the bonds that would be needed in place of this monitoring to keep shirking in check would be very large.

Another reason might be that bond/fee arrangements, as well as seniority wage arrangements, give employers a strong incentive to give their older workers poor evaluation reports; by firing more senior workers and replacing them with junior workers, they can either pocket the performance bond or fee, or avoid the payment of seniority bonuses.

Employer dishonesty might be kept in check by outside auditors (or even the legal courts), but this is rarely resorted to in practice. More important might be reputational effects. Firms which gain a reputation for dishonest treatment of workers often have trouble attracting good workers, hence reputation effects may encourage firms to deal honestly with their workers.

Clearly, much further work still needs to explain the general absence of sophisticated

contracts in labor markets.

References and Other Readings Related to Efficiency Wage Theory:

G. A. Akerloff (1984), *An Economic Theorist's Book of Tales* Cambridge University Press, Cambridge, UK.

G. A. Akerloff and J. Yellen, eds. (1986), *Efficiency Wage Models of the Labor Market*, Cambridge University Press, Cambridge, UK.

H. Dixon and N. Rankin, eds. (1995), *The New Macroeconomics: Imperfect Markets and Policy Effectiveness*, Cambridge University Press, Cambridge, UK.

G. Feiwel, ed. (1989), *The Economics of Imperfect Competition and Employment: Joan Robinson and Beyond*, New York University Press, N.Y.

R. Freeman (1998), "War of the Models: Which Labour Market Institutions for the 21st Century?," *Labour Economics* 5, 1–24.

P. Howitt (1990), *The Keynesian Recovery and Other Essays*, The University of Michigan Press, Ann Arbor.

A. Lindbeck and D. J. Snower (1988), *The Insider-Outsider Theory*, The MIT Press, Cambridge, MA.

N. Gregory Mankiw and D. Romer, eds. (1991), *New Keynesian Economics*, Vol.1: Imperfect Competition and Sticky Prices, and Vol. 2: Coordination Failures and Real Rigidities,

MIT Press Readings in Economics, Cambridge, MA.

K. M. Murphy and R. H. Topel (1990), "Efficiency Wages Reconsidered: Theory and Evidence," pp. 204-240 in: Yoram Weiss and Gideon Fishelson (eds.), *Advances in the Theory and Measurement of Unemployment*, Macmillan, London, 1990. [Theoretical and empirical critique of the New Keynesian theory of the efficiency wage.]

R. M. Solow (1979), "Another Possible Source of Wage Stickiness," reprinted as pp. 41-44 in Akerloff and Yellen, *op. cit.* [Classic article.]

J. Stiglitz (1992), "Capital Markets and Economic Fluctuations in Capitalist Economies," *European Economic Review* 36.2/3, 269-306. [An interesting readable overview of theoretical and empirical work on New Keynesian economics from 1980 through 1992 by one of the principal developers of the approach.]

O. E. Williamson (2000), "The New Institutional Economics: Taking Stock, Looking Ahead," *Journal of Economic Literature* XXXVIII, pp. 595-613.

Symposium (1997): Wage Inequality, *Journal of Economic Perspectives*, Spring 1997, Volume 11, Number 2. Articles by P. Gottschalk, G. E. Johnson, R. H. Topel, and N. M. Fortin and T. Lemieux.

Symposium (1993): Keynesian Economics Today, *Journal of Economic Perspectives* Winter 1993, Volume 7, Number 1. Articles by N. Gregory Mankiw, David Romer, James Tobin, and Robert G. King.