Working Paper 89-5/Document de travail 89-5

Efficiency Wages: A Literature Survey

by Timothy C.G. Fisher

Bank of Canada Banque du Canada

Efficiency Wages: A Literature Survey

Timothy C.G. Fisher
Research Department
Bank of Canada
Ottawa, Ontario
K1A 0G9

December 1989

The views expressed in this paper are those of the author; no responsibility for them should be attributed to the Bank of Canada. The author would like to thank Serena Ng and David Rose for helpful comments. Remaining errors or omissions are the sole responsibility of the author.

Abstract

Recently developed models based on a concept known as the *efficiency wage hypothesis* claim to have provided an explanation of persistent unemployment and rigid wages. The hypothesis states that labour productivity depends positively on the wage rate. If this is the case, firms may not reduce wages, even in the presence of unemployment, since reducing wages may actually increase labour costs through decreased productivity. This paper reviews both the theoretical models explaining a positive relationship between wages and productivity and empirical tests of the theory. Several literature reviews of the efficiency wage theory are currently available, but this survey focuses attention on the macroeconomic implications of the theory and reviews, for the first time, many empirical studies based on efficiency wage theory.

Résumé

Des modèles économiques élaborés ces derniers temps sur la base d'un concept connu sous le nom d'*hypothèse du salaire d'efficience* sont réputés être parvenus à expliquer la persistance du chômage et la rigidité du salaire. Cette hypothèse veut que la productivité du travail soit une fonction croissante du taux salarial. Partant de ce principe, les entreprises ne réduiront pas nécessairement le salaire même en présence de chômage, car la productivité pourrait se ressentir plus que proportionnellement d'une telle diminution, ce qui se solderait par un accroissement des coûts de main-d'oeuvre. La présente étude passe en revue les modèles théoriques qui postulent une relation positive entre le salaire et la productivité ainsi que différents tests empiriques des résultats prédits par la théorie du salaire d'efficience. Plusieurs études ont déjà brossé un tableau de la littérature portant sur cette théorie. Cette étude se distingue des précédentes en ce sens qu'elle s'intéresse surtout aux implications de la théorie au niveau macroéconomique et fait le tour pour la première fois de nombreux travaux empiriques fondés sur la théorie du salaire d'efficience.

Contents

Abstract/Résumé

Contents

List of Figures

1 Introduction 1

2 The Basic Model 2
2.1 Comparison to the competitive model 7
2.2 Extensions to the basic model 8

3 Microeconomic Foundations 9
3.1 The nutritional model 10
3.2 The shirking model 11
3.3 The labour turnover model 19
3.4 The adverse selection model 21
3.5 The sociological model 22
3.6 Summary 23

4 Macroeconomic Implications 24
4.1 Dual labour markets and interindustry wage differentials . . . 24
4.2 Unemployment 27
4.3 Cyclical fluctuations 28
4.4 Cyclical behaviour of labour markets 31
4.5 Wage rigidity 32

5 Further Criticisms 33

6 Empirical Tests 37
6.1 Evidence for a wage-productivity link 37
6.2 Evidence for interindustry wage differentials 40

7 Conclusion 45

References 48

List of Figures

1 Determination of the efficiency wage w^* 5
2 Alternative $b(w)$ function 6
3 Unemployment at the efficiency wage w^* 8
4 Equilibrium in the shirking model 13
5 The effect of the introduction of an unemployment benefit . . 14
6 The effect of a reduction in the marginal product of labour . . 29

1 Introduction

A satisfactory theoretical explanation of persistent unemployment has been elusive since Keynes first focused attention on the problem. On the one hand, it is not clear why unemployed workers do not offer to work at less than prevailing wages. On the other hand, even if such under-bidding is present, empirical observation suggests that firms do not reduce wages in the presence of unemployment. Recently, a class of theoretical models based on a concept known as *efficiency wages* claims to have uncovered an explanation for the absence of under-bidding by the unemployed and reasons why firms may not reduce wages when faced with an excess supply of labour. Hence efficiency wage models claim to provide a theoretical explanation for persistent unemployment. The present paper is a review of the literature on efficiency wages.

The basic idea behind efficiency wage theory is simple. The efficiency wage hypothesis states that worker productivity depends positively on the wage rate. This being the case, firms may not reduce wages, even when there is an excess supply of labour, since lowering wages may decrease productivity enough to increase total labour costs. Thus efficiency wage theory potentially explains persistent unemployment. Suitably adapted, efficiency wage theory potentially explains other observed yet previously unexplained features of labour markets such as rigid wages, variation in wages for similarly qualified workers across industries and variation in unemployment rates across distinct demographic groups in the labour force.

It is useful to divide the literature on efficiency wages into its theoretical and empirical components. The substantial theoretical literature on efficiency wages is essentially concerned with establishing a positive relationship between wages and productivity. The empirical literature on efficiency wages focuses on evaluating evidence for the wage-productivity link. Sections 2 to 5 concentrate on the theory of efficiency wages. To introduce key concepts,

section 2 examines the optimal input decision of a firm under the assumption of a positive relationship between the wage level and labour productivity. Five explanations have been offered for the efficiency wage hypothesis. Section 3 discusses each of these models of efficiency wages in turn. One of the most serious theoretical objections to some of the efficiency wage models, namely the so-called bonding critique, is also introduced in section 3. The implications of efficiency wage theory for rigid wages, cyclical fluctuations, unemployment and wage patterns are discussed in section 4. Theoretical and empirical objections concerning models of efficiency wages are raised in section 5. The empirical literature examining efficiency wage theory is reviewed in section 6. Lastly, section 7 contains some concluding remarks.

Several reviews of efficiency wage literature have recently been published: Yellen (1984), Stiglitz (1986), Akerlof and Yellen (1986), Katz (1986) and Carmichael (1987). The present survey differs from these papers in two important ways. First, it pays greater attention to the macroeconomic implications of efficiency wage theory. This is especially useful to readers interested in the implications of efficiency wages for macroeconomic policy issues. Second, most of the empirical studies of efficiency wages have been carried out since the last of the above papers was published, so the empirical efficiency wage literature is reviewed here for the first time. Ultimately, of course, the applicability of efficiency wage theory will depend on the degree to which it is able to explain observed economic behaviour.

2 The Basic Model

The basic efficiency wage model simply incorporates the assumption that wages and worker productivity are positively related into a rudimentary model of the firm. The firm is assumed to produce output, y, from two inputs: labour, L, and capital, K. Units of labour are assumed to be identical in every respect. The same assumption applies to units of capital. According

to the efficiency wage hypothesis, the actual amount of labour available to the firm for production is not simply the number of labour units, L, but rather the number of *efficiency labour units*, L^*. The number of efficiency units of labour depends both on the number of labour units hired and on the wage, w, paid by the firm. Formally this relationship may be written as follows

$$L^* = L^*(w, L). \tag{1}$$

Thus there are two types of labour in the model. First, there is the number of labour units, L. This is important to the firm since the total wage bill is measured by the wage rate times the number of workers hired, wL. Second, there is the number of efficiency labour units, L^*. This is important to the firm since total output is assumed to depend on the number of efficiency units of labour. As indicated by (1) the amount of efficiency labour depends on the wage level paid to the workers and on the total number of workers hired. For example, it is commonly argued in efficiency wage models that worker productivity or effort depends on the wage level.

In most applications of efficiency wage theory (1) is assumed to have a particular functional form[1]

$$L^* = b(w)L \tag{2}$$

where it is assumed that b is increasing in w, or $\partial b(w)/\partial w \equiv b_w(w) > 0$. In this case, $b(w)$ may be interpreted as the effort level of the workers. Output depends on capital and the number of efficiency units of labour, so the firm's production function is written

$$y = f(L^*, K), \tag{3}$$

or equivalently, using (2),

$$y = f(b(w)L, K). \tag{4}$$

[1]In the literature, the specific functional form for (2) is not justified because the basic efficiency wage model is used for expositional purposes only, as indeed it is here.

It is important to recognize the distinction between w, the wage rate per unit of labour, and the wage per efficiency unit of labour, $w/b(w)$. The latter represents the cost to the employer of purchasing an efficiency unit of labour and hence it is the relevant variable for factor demand decisions, given that output depends on efficiency units of labour. Suppose the firm can hire all the labour it needs at a wage no less than $\bar{w}$. At first glance it would seem that a rational firm would pay no more than $\bar{w}$ for labour, but this is not so. Since different wages purchase labour of different efficiency levels, the firm will choose to pay a wage that minimizes the cost per unit of efficiency labour, provided that the wage is at least as large as $\bar{w}$. In other words, the firm will choose w to minimize $w/b(w)$ subject to $w \geq \bar{w}$ regardless of the amount of labour hired. Once the wage is chosen, the firm will hire as many workers as are required to produce the optimal output level. Thus the firm's optimization decision involves two stages: (i) the firm chooses the wage level to minimize the average cost of an efficiency unit of labour $w/b(w)$ and (ii) the firm chooses the optimal level of output and inputs.[2]

The first stage of the firm's problem is formally written

$$\min_{w} \quad \frac{w}{b(w)} \qquad \text{subject to} \qquad w \geq \bar{w}. \tag{5}$$

Assuming the constraint does not bind, the first-order condition for the problem is

$$\frac{1}{b(w)} - \frac{w\, b_w(w)}{b(w)^2} = 0 \tag{6}$$

Multiplying through by $b(w)$ and rearranging, the first-order condition may be written

$$\frac{w\, b_w(w)}{b(w)} = 1 \tag{7}$$

The wage that solves this equation is called the efficiency wage and is denoted w^*. Equation (7) states that, at the efficiency wage, the elasticity of effort

[2]The two-stage optimization process is, of course, a direct result of the separability assumption present in (2).

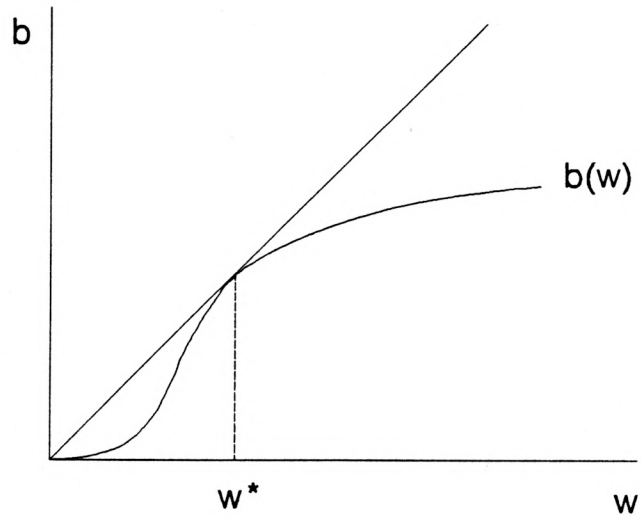

Figure 1: Determination of the efficiency wage w^*

with respect to the efficiency wage is unity. Equivalently (7) may be written

$$\frac{w}{b(w)} = \frac{1}{b_w(w)} \tag{8}$$

which states that the average cost of an efficiency unit of labour is equal to the marginal cost, $1/b_w(w)$, at the optimum wage. This, of course, is exactly as one would expect, since the firm is minimizing average cost of efficiency labour. The solution to the firm's wage problem may be illustrated, as shown in Figure 1. Rewriting (7) again as $b(w)/w = b_w(w)$, it follows that the efficiency wage is determined where the slope of the ray from the origin to a point on the $b(w)$ function is equal to the slope of the function at that point. In other words, the efficiency wage is determined at the tangent from the origin to the $b(w)$ function.

The wage-productivity relationship $b(w)$ is shown in Figure 1 as having a region of strict convexity and a region of strict concavity. Thus there are regions over which the wage-productivity relationship is subject to increas-

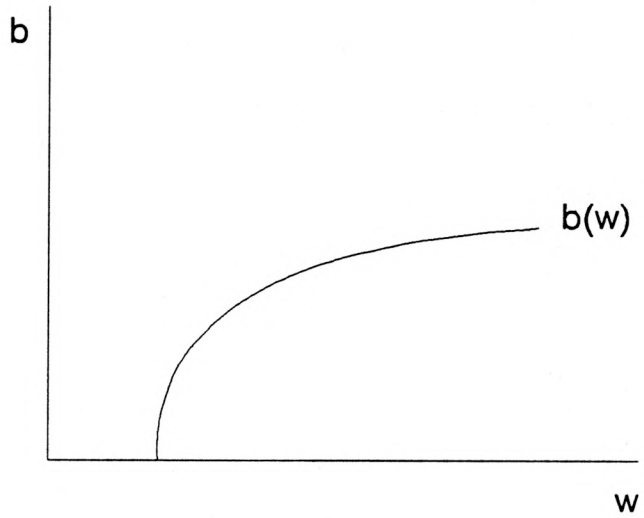

Figure 2: Alternative $b(w)$ function

ing returns, and regions over which the relationship is subject to decreasing returns. The region of convexity is not an essential part of the argument. However, if there is no region of convexity, the $b(w)$ function must intercept the horizontal axis to the right of the origin, as illustrated in Figure 2. In this case, there is a positive wage level below which there is no effort forthcoming from the workers. For a solution to the firm's wage problem to exist the $b(w)$ function must have one of the two shapes illustrated in Figures 1 and 2

Assume that capital is purchased in a competitive market at a per-unit rental price given by r. The second stage of the firm's optimization problem is to choose the optimal level of output, y, and the optimum levels of K and L given the efficiency wage, w^*, and the price of capital, r. Formally this is written

$$\max_{L,K} \quad f(b(w^*)L, K) - w^*L - rK \tag{9}$$

where the price of output is suppressed. The first-order conditions to this

problem are simply

$$\begin{aligned} f_L \cdot b(w^*) - w^* &= 0 \\ f_K - r &= 0 \end{aligned} \tag{10}$$

which are the familiar conditions that each optimal input level is determined where the marginal product is equal to the input price.[3]

2.1 Comparison to the competitive model

There are two essential differences between the basic efficiency wage model and the conventional competitive model of production. First is the presumption in the efficiency wage model that the effort or productivity of labour is sensitive to the wage level. In the standard competitive model, labour effort is implicitly assumed to be constant and, therefore, independent of the wage level. In terms of (1), the competitive model is simply represented by the identity $L^* \equiv L$. The second essential difference between the two models is that the firm is allowed to set the wage level in the efficiency wage model. In the competitive model, of course, the wage level is set in the market. It is these two departures from the competitive paradigm which are the essence of the efficiency wage approach and which yield the key implications of the model. Note that the two assumptions are related; it cannot be the case that the firm is paying efficiency wages and that the wage is set in the market at the same time.

Suppose that the market clearing wage is given by w^c. Then if the efficiency wage is greater than the market clearing wage, i.e. $w^* > w^c$, unemployment will be equal to the excess supply of labour at w^*, as illustrated in Figure 3. Note that the unemployment created by firms paying efficiency wages is involuntary; at the wage w^* there is an excess supply of workers willing to work at less than the going wage. Furthermore, firms are not willing

[3]Of course, unique input levels can only be solved from (10) if the technology is subject to non-constant returns to scale.

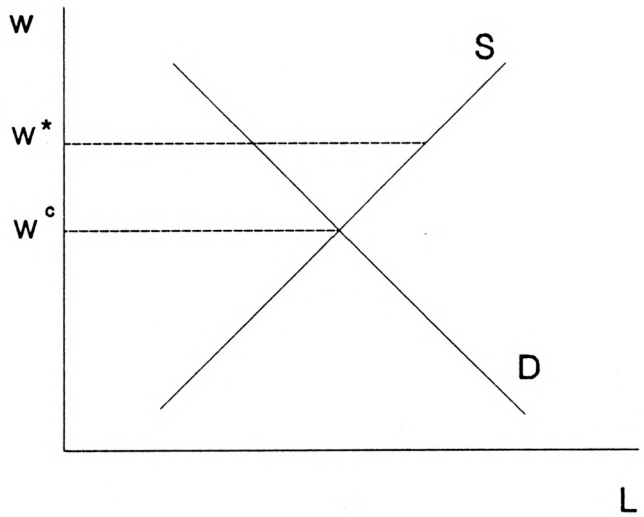

Figure 3: Unemployment at the efficiency wage w^*

to reduce the wage for fear of the adverse effect this would have on the productivity of the workers currently employed. Thus the efficiency wage model potentially explains both involuntary unemployment and rigid real wages. In a competitive labour market, of course, involuntary unemployment is ruled out since the wage adjusts to equate the supply and demand for labour. Two key features of Keynesian economics are the rigidity of wages and the existence of involuntary unemployment. The apparent ease with which efficiency wage theory explains these phenomena in turn explains the intense research activity the theory is currently generating.

2.2 Extensions to the basic model

In addition to rigid wages and involuntary unemployment, the efficiency wage model can be extended to explain several other observed or hypothesized labour market phenomena. One such hypothesized phenomenon is the existence of dual labour markets: the primary labour market, which is character-

ized by high paying, desirable jobs; and the secondary labour market, which is characterized by low paying, menial jobs. It could be argued that the primary sector is characterized by a strong wage-productivity relationship, and that the secondary labour market is characterized by a weak or nonexistent wage-productivity relationship. Given these assumptions, efficiency wage theory predicts the primary sector will have wages in excess of market clearing levels and a shortage of jobs, while the wage will be determined by competitive market forces in the secondary sector.

Alternatively, the link between wages and productivity may differ across firms in the economy. In other words, it may be that each firm in the economy has a different $b(w)$ function. Under these circumstances efficiency wages will differ from one firm to the next and, therefore, workers with identical characteristics may receive different wages in equilibrium. Such wage differentials are in addition to compensating differentials which arise to remunerate workers for welfare-reducing aspects of particular jobs. Finally, the relationship between wages and effort may vary across demographic groups in the labour force. That is, suppose each demographic group has its own distinct $b(w)$ function. In this case, each group will have its own efficiency wage and corresponding per-unit cost of efficiency labour. Firms will hire from the groups with the lowest efficiency labour cost. As a result unemployment rates will differ from one demographic group to another, rates being higher among groups with greater efficiency labour costs.

3 Microeconomic Foundations

The basic model presented above illustrates the essential features of efficiency wage theory. However, it is far from complete. A complete model must be able to explain the origins of the positive relationship between wages and productivity, rather than merely assume its existence. In other words the model must explain the presence of the $b(w)$ term in the production function

in (4).[4] In addition, it would be desirable if the model were able to explain some of the stylized facts about labour markets, such as: why firms use layoffs rather than wage reductions during periods of slack demand; why unemployed workers have characteristics similar to employed workers; and why unemployment is concentrated in distinct groups in the labour force.

The idea that labour productivity depends on the wage level can be traced as far back as Adam Smith,[5] while the term "efficiency wage" is credited to Hicks (1932). It is generally accepted, however, that Leibenstein (1957) represents the first formal efficiency wage model, and brief discussion of the Leibenstein paper is contained in section 3.1. The remainder of the section deals with the four other current explanations of the wage-productivity relationship. A brief summary is provided at the end of the section. It should be noted that some of the models described below are general equilibrium models and some are partial equilibrium models; care should be taken to distinguish the two. Second, without exception, all the models discussed below implicitly assume a closed economy. Thus the models do not take into account the effects of international trade or factor mobility.

3.1 The nutritional model

Leibenstein (1957) emphasizes the link between wages, nutrition and illness for agricultural workers in less developed countries. As agricultural workers in Third World countries live close to subsistence levels, the wage they receive is an important factor in determining their health. Higher wages permit better nutrition, thereby reducing the likelihood of disease. Since healthy workers are more productive, it follows that there is a positive relationship between wages and worker productivity. In richer, developed countries, however, the link between wages and worker health is not likely to be important because

[4] More generally, the model must explain the presence of the wage level in the production function.

[5] According to Eaton and White (1982).

wages far exceed subsistence levels. Consequently, the nutritional model of Leibenstein is not considered relevant to industrialized nations, although it has a place in explaining various features of labour markets in less developed countries, for example, the 'vicious cycle of poverty'—low consumption (wages) leading to low productivity leading to low consumption. For a formal development of the Leibenstein model, see Mirrlees (1976), Stiglitz (1976a) and Bliss and Stern (1978a).

3.2 The shirking model

The most common type of model in the literature on the efficiency wage hypothesis is the so-called shirking model. The argument typically proceeds as follows. Since workers find working distasteful, they will attempt to avoid it, or, to shirk. Workers caught shirking are fired, but since complete monitoring of workers by firms is not in general optimal, firms must find some other way to prevent shirking. One approach is to increase wages above the market clearing level. This increases the cost to workers of being caught shirking, and thus reduces the overall level of shirking. In effect, by raising the wage, firms purchase self-monitoring from workers.

Shapiro and Stiglitz (1984) construct a general equilibrium model where it is costly for firms to monitor the effort of individual workers. As monitoring is costly, it is not in the firms' interest to undertake complete monitoring. Workers have identical preferences over wages (which increase utility) and effort (which decreases utility). All firms have the same technology and behave competitively in the input and output markets. Workers who shirk have some probability of being caught, and if caught they are fired. In the standard competitive equilibrium to this economy, the wage adjusts to clear the labour market at the intersection of the labour supply curve and the aggregate labour demand curve. There is no unemployment in the competitive equilibrium. Zero unemployment implies that there is no penalty to shirking;

any worker who is fired for shirking can immediately be rehired. Therefore, with imperfect monitoring and full employment all workers will choose to shirk.

To induce workers not to shirk, each firm raises the wage above the market-clearing level. In this way, workers caught shirking pay a penalty. However, if it pays one firm to raise its wage, it pays all firms to raise their wages and the penalty to shirking is dissipated. Nevertheless, as all firms raise their wages, the aggregate quantity of labour demanded decreases and unemployment results. Now the worker who shirks is faced with the prospect of a spell of unemployment and there is an incentive not to shirk. That is, the presence of unemployment serves to discipline the workers. Moreover, unemployed workers are not able to bid for jobs by offering to work at a lower wage. Firms know that, at a lower wage, workers have an incentive to shirk, and the workers cannot make a credible promise not to shirk at a lower wage. Note that the actions of the firms are not co-ordinated; each firm raises the wage in an attempt to reduce shirking, but it is the collective result of their individual actions—unemployment—that eventually serves as the deterrent.

The Shapiro and Stiglitz model may be illustrated with the aid of a simple diagram of the labour market. The aggregate demand for labour, which is simply the horizontal sum of the individual firm labour demand curves, is shown in Figure 4. The wage level which is just sufficient to discourage shirking at each level of unemployment, known as the no-shirking wage, is depicted as an upward sloping curve, increasing with the level of employment. The positive slope of the no-shirking wage curve reflects the fact that, at a low level of unemployment the penalty to being caught shirking is reduced, so firms must pay higher wages to prevent shirking. Equilibrium in the aggregate labour market occurs where the wage level and the amount of unemployment are consistent with labour demand and no shirking, at w^* in Figure 4. If the wage is above w^*, workers value their jobs owing to

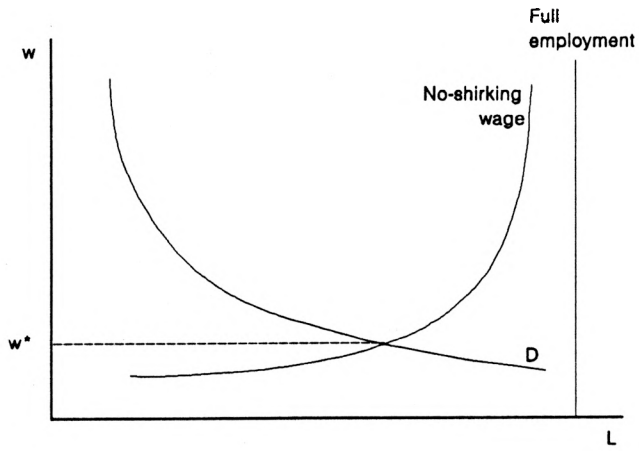

Figure 4: Equilibrium in the shirking model

the high wages and the low level of employment, meaning that firms can decrease the wage without creating the incentive to shirk. On the other hand, if the wage is below w^*, workers will not value their jobs because employment levels are high and wages are low, so firms must increase the wage to discourage shirking. Thus there are forces which tend to move the system toward equilibrium.

The Shapiro and Stiglitz shirking model has four important predictions. One of the most important implications of the model is involuntary unemployment. According to the model, the reservation wage[6] of the unemployed is strictly below the wage paid by firms to the employed. Since workers are identical, those without work are involuntarily unemployed; they would like to work, but there are no jobs available.

The second prediction of the Shapiro and Stiglitz model is that any vari-

[6]The reservation wage is defined as the wage level which is just sufficient to induce individuals to enter the labour force.

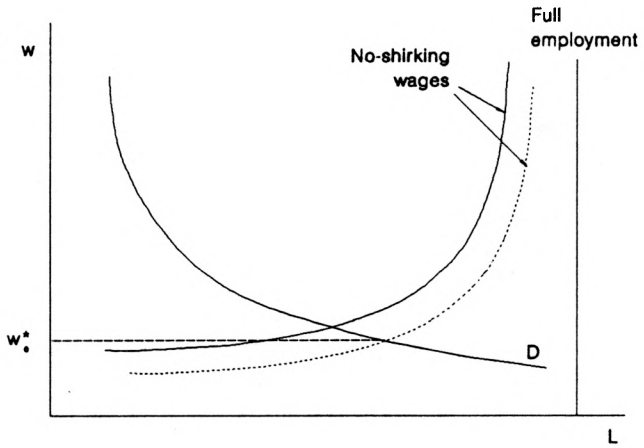

Figure 5: The effect of the introduction of an unemployment benefit

able which decreases the cost of unemployment to the worker, such as an unemployment benefit, will raise the equilibrium wage and increase the level of unemployment. The unemployment benefit reduces the penalty to shirking, so firms increase their wages in an attempt to restore the incentive not to shirk. In terms of Figure 5, the no-shirking wage increases at each level of unemployment following the introduction of an unemployment benefit and hence the no-shirking wage curve shifts to the left. As the wage increases to a new equilibrium level, aggregate labour demand decreases, unemployment rises and the incentive not to shirk is renewed.[7]

Thirdly, Shapiro and Stiglitz are able to show that the equilibrium level of unemployment in the shirking model is not Pareto efficient. That is, there exist wage subsidies which can make the firms and the workers better off.

[7]In this context, it is interesting to note that an empirical study by Weisskopf, Bowles and Gordon (1983) attributes some of the slowdown in U.S. productivity growth since the mid-1960s to decreases in worker effort brought about by increases in the unemployment benefit.

There are two types of externality present in the model. On the one hand, the private opportunity cost of hiring an additional worker is the wage rate, whereas the social opportunity cost is the effort level of the worker. Since in equilibrium the wage is greater than worker effort level, the private cost of employment exceeds the social cost which would lead to too few workers being hired. On the other hand, when a firm hires a worker the unemployment rate is decreased and, therefore, the wage paid by other firms must increase. Individual firms ignore this effect, which would lead to too many workers being hired. The former of these two effects dominates, implying that the equilibrium rate of unemployment is greater than the social optimum, and that the private cost of employment is larger than the social cost in equilibrium. A wage subsidy that reduces the private cost of employment until it is equal to the social cost will result in an efficient level of employment. Note that even in the absence of externalities, the Pareto-efficient level of unemployment is not zero because unemployment performs the socially useful task of preventing shirking.

Lastly, if the assumption of identical firms is relaxed, it is possible for the Shapiro and Stiglitz model to generate wage distributions for identical workers. For example, the costs associated with shirking may differ from one firm to the next, or monitoring costs may differ across firms. Firms with greater monitoring costs will have a lower level of monitoring and, consequently, higher wages to discourage shirking relative to firms with lower costs of monitoring. Therefore, otherwise identical workers will receive different wages in equilibrium.

Foster and Wan (1984) interpret the issue of employee shirking as a principal-agent problem. According to their view, the firm (principal) is unable to observe the effort of the worker (agent), but the firm does know the distribution from which effort is drawn. The firm's problem is to choose the number of workers and a contract for each worker to maximize expected profit subject to the unobservability of worker effort. As is typical in models

with asymmetric information, the equilibrium in the Foster and Wan model is second-best; the optimal contract requires unemployment in order to induce the right amount of effort from the workers. Foster and Wan demonstrate that the equilibrium unemployment in their model is involuntary in the sense that employed workers enjoy a strictly higher level of expected utility than unemployed workers.

Bowles (1985) views the employee shirking problem as a microcosm of an issue raised by the eighteenth century English philosopher Hobbes concerning the reconciliation of individual self-interest and collective or societal rationality. The contemporary issue is how to resolve the conflict of interest between the worker, who dislikes work, and the employer, who requires work. Bowles calls the shirking problem a "neo-Hobbesian" issue. Although the spirit of Bowles' approach is quite different from that of Shapiro and Stiglitz, the end result is the same: in equilibrium, the threat of unemployment acts as the incentive to encourage the right amount of effort from the workers. There are numerous other models examining the shirking problem. See, for example, Eaton and White (1982) and Miyazaki (1984). Although these models differ depending on what is observable, at what cost, and the nature of feasible contracts, the models share an assumption that the firm pays a fixed wage to the workers.

The major theoretical objection to the shirking models is that more sophisticated compensation schemes can be used by a firm to eliminate the shirking problem. According to this objection it is not necessary for unemployment to act as the incentive to discourage shirking and, furthermore, involuntary unemployment may be eliminated altogether in equilibrium. This objection is usually called the *bonding critique*. For example, consider an employment fee paid by the worker to the firm upon starting work. If all firms charge employment fees, workers will have an incentive not to shirk since workers caught shirking will have to pay another fee to become employed again. The introduction of employment fees also reduces the cost of labour

to firms, so more workers will be employed. In equilibrium, the employment fee will be such that the expected value of wages to the worker is equal to the expected value of alternative activity such as home production. That is, involuntary unemployment is eliminated in equilibrium. The employment fee substitutes for the threat of unemployment as the incentive not to shirk. A scheme with similar results requires workers to post a bond when they join the firm. If the worker is caught shirking, the bond is forfeited and the worker has to post another bond to become re-employed. A final example is for the firm to design an upward-sloping age-earnings profile to compensate workers. Newly hired workers are paid less than their marginal product, and older workers are paid in excess of their marginal product. Workers caught shirking forgo the wage premium later in their careers, providing the incentive not to shirk. The age-earnings profile will adjust in equilibrium to equate the present discounted value of jobs with the present value of the workers' alternatives (e.g. home production) and the involuntary nature of the unemployment disappears.

There are two responses to the bonding critique. First, Shapiro and Stiglitz (1984) point out that imperfect capital markets mean that workers may not have enough cash to post a bond or pay an employment fee. Second, fees and bonds paid by the workers give rise to a moral hazard problem on the part of the firm: once the bond is posted, the firm may simply claim the worker has shirked, cash the bond and fire the worker. Several extensions to the Shapiro and Stiglitz model have been proposed which argue that moral hazard on the part of the firm and capital market imperfections may not be severe. First, Carmichael (1985) points out that although the firm may not be able to charge an employment fee to equalize the expected *value* of lifetime wages with the expected value of the alternative, owing to capital market imperfections, the firm may be able to charge a fee which equates the expected lifetime *utility* of work and its alternative. Since a fee which is designed to affect the utility of work will be smaller than a fee designed

to affect the value of work, the capital market imperfections may not be as severe. Further, a fee which equates the lifetime utility of work with the lifetime utility of the alternative removes the involuntary nature of the unemployment. Second, Lazear (1981) points out that if the firm is worried about its reputation in the labour market, it will not be in the firm's interest to cash-in worker performance bonds when shirking has not occurred. Thus the moral hazard problem of the firm may not be severe. Finally, Carmichael (1983) and Malcomson (1984) offer alternative contract schemes which eliminate the moral hazard problem and the shirking problem provided there are perfect capital markets. In these models, the worker pays a performance bond into a pension fund. Shirking workers are fired and the bond defaults not to the firm but instead to a pension fund which is distributed among the remaining workers at the firm. Workers who do not shirk are eligible to reclaim their bond, with interest, upon retirement. As the firm cannot claim performance bonds there is no moral hazard problem. And the existence of bonds provides adequate incentive to workers not to shirk.

Shapiro and Stiglitz (1985) respond by noting that performance bonds are rarely observed in the real world, probably because of the moral hazard problem associated with the firm. Further, they take exception to Carmichael's definition of involuntary unemployment, i.e. that if worker utility from employment is equal to the utility from the alternative then the unemployment is voluntary. They paraphrase Carmichael's definition as follows: as long as work is available somewhere in the economy—for example, in the service sector—workers without jobs cannot be viewed as involuntarily unemployed. Shapiro and Stiglitz reiterate that the unemployment in their model, whether it is termed involuntary or otherwise, is shown to be inefficient and hence some form of intervention is warranted. A formal model which responds to the bonding critique is provided by Beaudry (1989), discussed below in the context of the turnover version of efficiency wage theory.

3.3 The labour turnover model

An alternative justification for the efficiency wage hypothesis is the labour turnover model presented in various forms by Phelps (1970), Stiglitz (1974) and Salop (1979). The basic hypothesis of the approach is that labour turnover is costly to the firm because new employees have to be trained owing to the existence of firm-specific human capital. Thus firms may attempt to increase their wage relative to other firms in an attempt to discourage workers from quitting. If firms are identical, all firms will raise their wages and relative wages will not change. Aggregate employment will, however, be reduced. Therefore, the end result is the same as the shirking model, except in this case equilibrium unemployment serves to reduce labour turnover rather than to decrease shirking. In this model the wage premium enhances the net productivity of workers by reducing turnover costs.

It is not unreasonable to suppose that turnover costs will differ between firms. In this case, firms with high costs of turnover will pay a larger wage premium over the market-clearing level. Hence, the turnover model may explain wage differentials for identical workers in much the same way as the shirking model. Given the similarity between the two, it is not surprising that the criticisms which apply to the shirking model apply to the turnover model as well. Thus one objection to the turnover model is that more sophisticated contracts can eliminate the need for unemployment to diminish turnover. As Salop points out, the problem facing the firm is that it is unable to pay different wages to newly hired and senior employees. Suppose the firm is permitted to pay new employees a wage equal to the difference between their marginal product and their training cost. Fully trained employees may then be offered a rising wage profile based on seniority. The incentive not to quit is provided by the rising wage profile. Rising wage profiles based on seniority are commonly observed. Furthermore, unlike the shirking model, there is no moral hazard problem on the part of the firm since the firm has no incentive

to fire trained workers.

Beaudry (1989) contains the first formal model to challenge the bonding critique. He introduces an information asymmetry into the turnover model: the firm is assumed to be better informed than the prospective employee about the amount of firm-specific human capital the employee will acquire. If hired, the employee will undergo a fixed training period and be paid a training wage. After the training period, the wage is renegotiated and the worker is paid a post-training wage for the remainder of the job tenure. Turnover is costly to the firm, and jobs are available outside the firm at some alternative wage. In deciding whether or not to work at the firm, the prospective employee or applicant would like to know the post-training wage. However, since only the firm knows the value of training, the applicant cannot infer the post-training wage. Further, a promise by the firm to pay a post-training wage in excess of the alternative wage available to the applicant is not credible: the firm could simply renege on the promise and pay the worker the alternative wage.

It turns out that the firm can credibly promise to pay a post-training wage above the alternative wage provided it pays *training wage* in excess of the alternative wage. Thus the firm offers a training wage which dominates the applicant's alternative in order to convince the applicant that the post-training wage will continue to pay the worker some rent. Put another way, the training wage is a signal of the credibility of proposed future wage payments. In turn, the credibility of the firm's post-training wage offer reduces turnover at the firm, thereby justifying wage payments in excess of the worker's alternative. Entrance fees or employment bonds are not used by the firm since they will be interpreted by the worker as a signal that the post-training wage will not be honoured, resulting in an increase in turnover.

In addition to the prediction that higher turnover costs imply higher training and post-training wages, the Beaudry turnover model has two other predictions. First, since wages at the firm are determined by turnover costs

at the firm, the wage profile at the job is independent of the observed wage change between the employee's previous job and the new job. Second, there is a linear relationship between training and post-training wages. That is, the training wage is a linear predictor of future or post-training wages with a coefficient of one. Beaudry tests the predictions of the model using individual wage data; his results are discussed in section 6.2.

3.4 The adverse selection model

Another potential explanation of the positive relationship between wages and productivity is the adverse selection model discussed in Stiglitz (1976b) and Weiss (1980). In this model, it is assumed that worker productivity depends on "ability" and that workers have different abilities. An important assumption is that workers with higher ability are assumed to have higher reservation wages. Asymmetry of information arises in the model as firms do not observe worker ability, nor are firms able to infer worker ability after the workers have been hired. Given the imperfect information, firms must hire workers randomly from the pool of job applicants. Suppose that, at a particular wage level, a firm is able to attract enough job applicants to satisfy labour demand. Then, given the assumption that worker reservation wages rise with ability, increasing the wage offer raises the expected ability level of a worker drawn randomly from the applicant pool. Hence, the firm pays a wage in excess of the market clearing level in order to secure a better pool of applicants. Thus, it is argued, the model explains the existence of job queues. Furthermore, a job applicant cannot obtain work by offering to work at less than the going wage since the offer signals to the firm that the individual is a low-ability worker.

If there is a decrease in demand at the firm, the firm may respond either by cutting wages or laying off workers. Since the firm is not able to observe individual worker ability, layoffs of the least productive workers are not fea-

sible. Cutting wages, on the other hand, results in the better workers leaving the firm. Thus the only option available to the firm in response to a decline in demand is to invoke random layoffs. Hence the model explains the existence of layoffs and the presence of inflexible wages. If identifiable groups in the labour force have different reservation wages, the adverse selection model predicts wage differentials and different layoff probabilities among the distinct groups.

Critics of the adverse selection approach focus on the assumption that the firm is unable to observe worker productivity even after the worker has been with the firm for some time.[8] It seems plausible that the firm eventually becomes aware of each worker's ability. This being the case, a bonding scheme identical to that described for the shirking model could be used by the firm to eliminate the adverse selection problem.

3.5 The sociological model

The above four types of rationalization of the efficiency wage hypothesis are strictly within the neoclassical paradigm. Another approach to the problem avoids the assumption of individual maximization and concentrates instead on a sociological explanation. Solow (1979) asserts that wage payments depend to some degree on social conventions and notions of fair treatment. For example, firms may not cut wages in times of excess labour supply for fear of adversely affecting worker morale, which in turn reduces productivity. In addition, it may not be in the firm's long-run interest to cut wages because it might gain a reputation as an unfair employer and have difficulty attracting workers in the future. Akerlof (1982) hypothesizes that worker productivity depends on the work norms of fellow employees. Firms can increase the group work norms of employees by paying wages in excess of the wage re-

[8] A model where the firm does observe worker ability is contained in Gibbons and Katz (1989). However, this model does not generate equilibrium unemployment, focusing instead on wage differentials for workers who have been laid off.

quired to yield a minimum level of effort from the workers. The excess wage payment is termed a "gift" by Akerlof. Thus in return for the gift of higher wages, workers give the gift of increased work effort. Finally, Bowles (1985) emphasizes the sociological aspects governing the worker effort decision in a Marxian model of extraction of labour power from labour.

3.6 Summary

The present section has examined five models which purport to explain the positive relationship between wages and productivity known as the efficiency wage hypothesis. It is argued that the nutritional model is of limited relevance to advanced capitalist economies. Either of the four other explanations for the efficiency wage hypothesis are potentially relevant to some degree or other. A consensus has not yet emerged on their relative merits. There is a need for further theoretical research—especially regarding the issues raised by the bonding critique—and empirical research.[9]

The challenge the bonding critique poses for efficiency wage theory is to explain why more complicated contracts, which remove the need for unemployment to "discipline the workers," are not used by firms to elicit the right amount of effort from their workers. A related issue is that complicated contracts are rarely observed in the labour market, and it would be useful if there were theories to explain this fact. At the moment, this question has invited some conjecture but little theoretical research. Akerlof and Yellen (1986) muse that complicated contracts are avoided by firms and workers because they necessitate comparisons between worker performance which workers would find distasteful. Another factor militating against the use of complicated contracts is that the opportunity for disagreement or misunderstanding between the contracting parties is likely to increase with

[9]Section 6 reviews existing empirical research and suggests ways to distinguish empirically between the various efficiency wage models.

the complexity of the contract. Shapiro and Stiglitz (1984) hypothesize that bonding-type contracts are rarely observed because of the moral hazard problem this presents to the firm.

4 Macroeconomic Implications

The efficiency wage models described above have interesting and important macroeconomic implications. Reflecting the focus in the literature, the discussion here concentrates on the macroeconomic implications of the shirking and turnover models with less emphasis on the implications of the adverse selection and sociological models. The nutrition version of efficiency wage theory is not considered further.

4.1 Dual labour markets and interindustry wage differentials

Institutional labour economists have often observed that the labour market is essentially composed of two sectors: a primary sector where wages are high and job tenure is long, and a secondary sector where wages are low and jobs are menial.[10] Doeringer and Piore (1971) introduced the concept of a *dual labour market* comprised of primary and secondary sectors, where workers in the primary sector enjoy a surplus relative to workers in the secondary sector, who achieve only a reservation level of utility, and where there is a shortage of primary sector jobs. Although the Doeringer and Piore model approximates observed labour markets, it is subject to a fundamental criticism: why is the surplus enjoyed by the primary sector workers not eventually bid away?

Recently Jones (1985) and Bulow and Summers (1986) have proposed models of dual labour markets where the high wages in the primary sector are not bid away, providing a rationale for persistent wage differentials

[10]Dickens and Lang (1985a, 1985b) construct a switching regression model of a dual labour market and find statistical evidence supporting its existence in the U.S.

between the primary and secondary sectors. The insight of this approach is to recognize that jobs in the two sectors are different. In the primary sector, jobs are long-lived and require a great deal of independent and unsupervised action on the part of the worker. In the secondary sector, jobs are short-lived, comparatively unskilled and easily supervised. As supervision is more difficult in the primary sector, firms in the primary sector may use alternative methods to ensure adequate performance from their employees. Jones (1985) shows that workers in the primary sector receive a wage premium analogous to the premium received by employed workers in the Shapiro and Stiglitz (1984) shirking model. The secondary sector behaves like a conventional competitive labour market, where the wage is bid down to the market-clearing level. The wage premium creates unemployment in the primary sector which serves to discipline the primary sector workers and compensate for the lack of supervision.[11] The wage premium in the primary sector will persist in equilibrium because primary sector firms have no incentive to lower the wage; to do so would only encourage shirking.

Of course, the primary-secondary division is purely arbitrary, there being a continuum of possible divisions. At one extreme, it is not implausible to suppose that job tenure and the level of supervision are different for every firm in the economy, and hence that in equilibrium there will be as many different wage levels as there are firms. More realistically, tenure and supervision may differ from one industry to the next. For example, monitoring costs and costs of shirking or turnover may be higher in some industries than in others. Accordingly, wages will differ across industries even though workers are identical, and queues may develop to regulate entry by workers into the higher paying jobs. Thus a straightforward extension of the assumptions of efficiency wage theory gives rise to wage differentials and differences in unemployment across industries.

[11]Some unemployed workers are essentially waiting for primary sector jobs, forgoing secondary sector employment in the hope of securing more lucrative work.

The sociological model predicts that efficiency wages are more likely to arise where teamwork and work in groups are important. If there are differences in the degree of group work or teamwork across industries, the sociological model predicts interindustry wage differentials for identical workers. Interindustry wage differentials are predicted by the adverse selection model, provided different industries require workers of various abilities. However, since industries paying higher wages will on average attract higher ability workers, once worker quality is controlled for, the interindustry wage differentials may not remain.[12] Thus the adverse selection model does not appear capable of explaining wage differentials for identical workers.

In the standard competitive model, the equilibrium wage will be the same in all industries once allowances are made for labour quality and compensating differentials.[13] Any industry where the wage exceeds the norm will attract more workers, causing the wage to fall until it is the same as in other industries. Hence, the existence of significant interindustry wage differences, controlling for compensating differentials and differences in labour quality, is an important indication that the standard competitive model is not applicable. In contrast, efficiency wage theory provides a potential explanation for interindustry wage differentials. Nevertheless, evidence of wage differentials is neither a necessary nor a sufficient condition for the existence of efficiency wage payments. For example, wage differentials may arise for institutional reasons or other factors which have nothing to do with efficiency wages. Alternatively, the absence of wage differentials may simply be due to similar efficiency wage payments in all firms.[14]

[12]This presumes that worker quality can be controlled for when, according to an assumption of the adverse selection model, it is unobservable.

[13]A *compensating differential* is a wage premium paid to compensate workers in a particular industry for dangerous or unpleasant work. For example, an unskilled labourer working for a city sanitation department may receive a higher wage than an identically qualified worker at the Bank of Canada. In this example, the sanitation worker receives a compensating differential.

[14]For a discussion of empirical evidence for interindustry wage differentials see sec-

4.2 Unemployment

The existence of equilibrium unemployment in models of efficiency wages raises two questions. First, is the equilibrium level of unemployment to be characterized as voluntary—i.e. at the discretion of the worker—or involuntary? Second, what, if anything, do the efficiency wage models have to say about the natural rate of unemployment, i.e. the long-run equilibrium level of unemployment? With respect to the second question, it is clear that the unemployment in the shirking and turnover models is natural unemployment. Equilibrium unemployment arises in these models as a result of attempts by firms to reduce costly shirking or turnover by workers. As such costs persist in the long run, unemployment in the shirking and turnover models is in fact long-run equilibrium or natural unemployment.[15] This suggests that changes in the natural rate of unemployment may be linked to changes in turnover or shirking costs over time.

In the Shapiro and Stiglitz (1984) shirking model and in the turnover model of Salop (1979), the equilibrium unemployment is clearly involuntary unemployment. Unemployed workers strictly prefer to work at a wage less than the efficiency wage offered by firms but there are no jobs available. However, when the dual labour market version of efficiency wage theory is considered, it is not obvious that workers without jobs are involuntarily unemployed. Since there are always jobs available in the secondary sector, albeit low paying jobs, unemployed workers in the primary sector are, in one sense, voluntarily unemployed. On the other hand, since the unemployed workers and those working in the secondary sector are strictly worse off than the primary sector workers and since all workers are otherwise identical, those without work are, in another sense, involuntarily unemployed. As an example, the erstwhile manufacturing worker who turns down a job sweeping

tion 6.2.

[15]Thus the aggregate supply curve in the shirking and turnover models is vertical at the natural or long-run equilibrium level of unemployment.

floors in a fast-food restaurant is voluntarily unemployed in the former sense and involuntarily unemployed in the latter sense. Jones (1985) argues that, to the extent that policy is concerned with the welfare of individuals, the focus of interest for policy actions is the availability of good primary-sector employment opportunities.

Lastly, to the extent that the relationship between the wage level and productivity differs from one group in the labour force to the next, efficiency wage theory predicts different unemployment rates across the groups. In addition, changes in aggregate demand will have different effects on the employment levels of the various groups. Both these predictions are consistent with the observed concentration of unemployment among various groups in the labour force, such as younger workers, and the changes in relative unemployment rates across groups over the business cycle.

4.3 Cyclical fluctuations

Although efficiency wage models have little, if anything, to say about the causes of cyclical fluctuations in aggregate output, they do potentially offer some insight into the propagation mechanisms of external shocks. Since any model based on pure maximization is necessarily a real model, the neoclassical efficiency wage models[16] are clearly real wage models and hence equilibrium is neutral. That is, if all exogenous nominal variables change proportionately, then the equilibrium set of endogenous variables must change in the same proportion, leaving the equilibrium set of real variables unchanged. Accordingly, equilibrium unemployment in efficiency wage models cannot be affected by changes in the money supply. Furthermore, as noted above, the aggregate supply curve in the shirking and turnover models is vertical at the long-run equilibrium level of unemployment, so any fluctuations in aggregate demand will have no effect on output and employment. Thus it may

[16]That is, all the efficiency wage models except the sociological model.

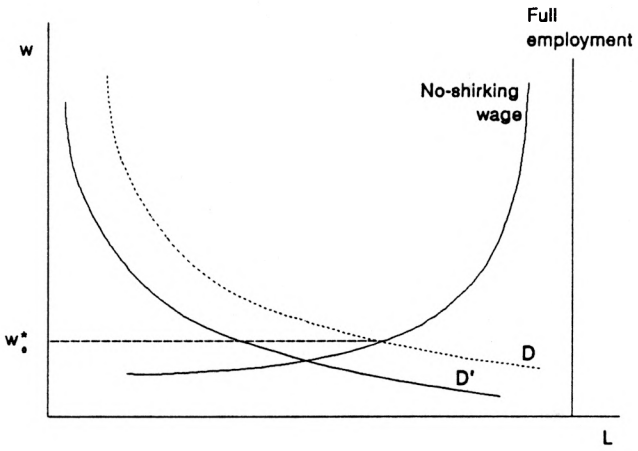

Figure 6: The effect of a reduction in the marginal product of labour

seem that efficiency wage models are unable to explain even the propagation mechanisms of shocks. In fact, the shirking and turnover models must be enriched if they are to explain the propagation of cyclical fluctuations.

Two ways of extending efficiency wage models to explain the propagation of cyclical fluctuations are proposed in the literature. Shapiro and Stiglitz (1984) emphasize the role of *co-ordination problems* in the adjustment of the economy from one equilibrium to another. To illustrate, consider an economy-wide shock which reduces the demand for labour at each firm. Assume that each firm takes the behaviour of all other firms to be fixed.[17] A pre-shock equilibrium wage of w_0^* is shown in Figure 6. The reduction in the marginal product of labour at each firm shifts the aggregate demand for labour to the left. At the pre-shock level of employment, w_0^* exceeds the post-shock marginal product of labour at each firm. Since each firm takes

[17]Equivalently, it could be assumed that firms are unable to distinguish economy-wide demand shocks from firm-specific shocks.

the behaviour of other firms as fixed, individual firms have no incentive to cut wages. Each firm believes that if it cuts wages, its workers would have an incentive to shirk since they could find work elsewhere at a higher wage. Layoffs are therefore the optimal response to the shock for each firm. Only when it becomes apparent to all firms that the aggregate unemployment level is increasing will individual firms begin to reduce their wages without fear of inducing shirking. Thus the lack of co-ordination of firms' wage-setting decisions results in inertia in the wage-adjustment process and overshooting of the new equilibrium level of unemployment.

Summers (1988) also emphasizes the role of co-ordination problems in preventing the immediate attainment of equilibrium following a shock. He points out that for adjustment to be instantaneous, all firms must simultaneously be aware of the wages being paid elsewhere and that firms must be able to distinguish economy-wide shocks from firm-specific declines in demand. Expecting firms to meet such informational requirements is unrealistic, claims Summers, giving rise to co-ordination problems. Stiglitz (1986) considers the case of staggered wage contracts, where contracts are renewed at intervals by firms. Since the optimal wage at one firm depends on the wages at other firms and on the level of unemployment, if some firms do not adjust, it will not pay the remaining firms to adjust as rapidly as they otherwise might. Hence adjustment between long-run equilibrium levels will occur slowly.

Although the introduction of co-ordination problems appears to imbue efficiency wage models with the ability to explain propagation of economy-wide shocks, co-ordination problems *per se* are not unique to efficiency wage theory. Appending co-ordination problems to a conventional competitive model of the labour market will also induce slow adjustment between equilibriums. However, a second explanation for the slow attainment of equilibriums following shocks is unique to efficiency wage theory. In any model of efficiency wages, the wage is chosen optimally by the firm. Thus it follows from the

envelope theorem[18] that firms that fail to adjust the wage in response to small changes in exogenous variables will suffer losses of only a second-order magnitude. In contrast, the standard competitive paradigm predicts that firms that fail to adjust wages in response to the market will suffer losses of a first-order magnitude. Since firms paying efficiency wages suffer only second-order losses, firms failing to choose wages precisely optimally following a shock face small losses. Such *near-rational* behaviour can give rise to first-order employment and output effects in response to shocks, as shown by Akerlof and Yellen (1985b).

Monetary policy may also affect aggregate demand in an efficiency wage model. An increase in the real rate of interest reduces the present value of a job to a primary sector worker, increasing the incentive of such workers to shirk. To restore the incentive not to shirk, primary sector firms would increase the wage, increasing the equilibrium level of unemployment. Although this link between monetary policy and unemployment is theoretically plausible, it is unlikely to be of much empirical relevance.

4.4 Cyclical behaviour of labour markets

According to the shirking model, a rise in the unemployment rate and consequent lengthening in unemployment spells increases the cost to workers of quitting jobs and this in turn implies fewer quits. Thus the shirking model predicts procyclical quit rates. In contrast, most search models of unemployment predict more quits when unemployment rises.

A well established stylized fact concerning the behaviour of labour markets is that most of the cyclical variation in labour is concentrated in employment fluctuations rather than in variations in hours of work per employee. The adverse selection model explains the preponderance of layoffs in response to cyclical downturns by noting that wage cuts and work sharing cause the

[18]For a succinct exposition of the envelope theorem, see Varian (1978), pages 267–268.

more able workers to leave the firm. Layoffs are the only alternative for firms that wish to retain their better workers and at the same time reduce their labour demand. An alternative reason for the use of layoffs rather than work sharing in response to declines in demand is provided by the shirking model of Bulow and Summers (1986). Bulow and Summers argue that the value of a job is greater to a full-time worker than to a part-time worker. Work sharing, i.e. changing full-time workers into part-time workers, reduces the value of jobs to workers, necessitating increased wages to discourage shirking. Layoffs, on the other hand, preserve job value for the remaining full-time employees and hence maintain work incentives without the need for wage increases. However, as Katz (1986) points out, the argument of Bulow and Summers is more applicable to permanent declines in demand rather than transitory or cyclical declines, as the value of a job to a worker is a long-term concept. Lastly, as noted above, layoffs will be used by firms in response to cyclical decreases in demand in the presence of co-ordination problems.

4.5 Wage rigidity

There are two aspects of efficiency wage theory that could help explain the rigidity or stickiness of wages. First, it is clear that, since equilibrium unemployment is a feature of all efficiency wage models, wages do not respond to clear the market in these models. In this sense, rigid real wages are an implication of all efficiency wage models. The turnover and the sociological models may be used to explain, albeit not very convincingly, the existence of rigid nominal wages. In the sociological model, if workers believe that money wage reductions are unfair, it is in the interest of firms to maintain the level of money wages. Even though the behaviour of the workers is irrational, it pays the firm to reflect irrationality in their wage setting. This explanation may be plausible in the short run for an economy which has had little experience with inflation, but it is not a convincing explanation for the long

run.

The second aspect of sticky wages concerns the adjustment of wages from one equilibrium to another following an external shock. Again, as stated above, Shapiro and Stiglitz (1984), Stiglitz (1986) and Summers (1988) claim that the lack of co-ordination of firm wage decisions means that the adjustment of the economy to an external shock will be a slow process. Thus in the turnover model, a change in the demand or supply of labour will leave the equilibrium wage unchanged if each firm believes that all the other firms are not going to change their wage. The equilibrium wage could be defined in money terms, perhaps in an economy which has little experience with inflation, or it could be defined in real terms for economies where indexing is more common.

5 Further Criticisms

It is probably safe to say that the most serious theoretical objection to efficiency wage theory is the bonding critique discussed in section 3. To reiterate, the bonding critique states that efficiency wage payments to ensure adequate worker performance are unnecessary because firms have alternative disciplinary schemes at their disposal. Essentially these schemes force the worker to post an implicit performance bond which is forfeited if the worker does not perform satisfactorily. Examples are seniority wage systems and pension schemes. In this context, there is an additional point worth noting. The dual labour market models of Jones (1985) and Bulow and Summers (1986) predict that efficiency wage considerations are more relevant in the primary sector of the labour market. But the primary sector is precisely where bonding schemes, are likely to be more prevalent, owing to the long tenure of primary sector jobs. Thus the need for efficiency wage payments in the primary sector is mitigated. Indeed some have argued that efficiency wage considerations are more important in dead-end, i.e. secondary sector,

jobs because such jobs provide no room for advancement or other ways of rewarding satisfactory employee performance.[19] Ultimately, whether seniority wage systems and pension schemes provide full bonding or whether there is a role for efficiency wages is an empirical question.

Consider again the firm's wage optimization problem in the basic efficiency wage model of section 2. The first-order condition for the problem is arranged to show that the elasticity of effort with respect to the optimal efficiency wage is unity.[20] In other words, the equilibrium wage is such that, for example, a 10 per cent increase in the wage will induce a 10 per cent increase in worker effort. A potential criticism of efficiency wage theory is that this elasticity is implausibly high. Akerlof and Yellen (1986) present a convincing example, in a slightly modified version of the basic efficiency wage model, where the elasticity of effort with respect to the wage is less than unity. In their example, shirking reduces the firm's output directly by decreasing the amount of efficiency labour input and indirectly by wasting or misusing other inputs. Compared to the basic model, a high level of effort is more important in the Akerlof and Yellen model and, therefore, the equilibrium effort-wage elasticity is lower. The applicability of the Akerlof and Yellen example depends upon how efficiency labour is assumed to interact with other factors of production, which raises a related criticism of efficiency wage theory: the functional form of the production function.

In the basic efficiency wage model, the production function is assumed to take the form

$$y = f(b(w)L, K), \tag{11}$$

where y is output, w is the wage, L is labour input, K is capital input, $b(w)$ is the effort function, and $b(w)L$ is efficiency labour. In the above specification, the wage enters the production function in a labour-augmenting

[19] For example, see Weiss (1986).

[20] See equation (7) in section 2.

way. Consider the general form of the production function

$$y = f(w, L, K). \tag{12}$$

Writing the production function as (11) rather than as (12) implicitly embodies two assumptions. First, it assumes that changes in effort, brought about by changes in the wage rate, result in proportionate changes in the amount of efficiency labour. Thus, for example, in terms of units of efficiency labour, a 5 per cent increase in effort is equivalent to a 5 per cent increase in the number of workers. Second, it assumes that a change in effort affects only the efficiency labour input; all other inputs remain unaffected. As Bliss and Stern (1978a) point out, in general this cannot be correct. Consider two modes of production: one a highly automated, capital-intensive process, the other a primitive, labour-intensive process. It is unlikely that effort and the number workers will interact in exactly the same way to form efficiency labour in the two processes. In other words, it is unlikely that a restrictive specification, such as the one represented by (11), will suitably model production in a large cross-section of firms or industries.

Solow (1979) shows that the particular specification of the production function in (11) is necessary and sufficient for complete (real) wage rigidity in the basic efficiency wage model. That is, the firm will not change the wage in response to changes in output if and only if the $b(w)$ term enters the production function in the labour-augmenting manner shown in (11). In view of the Bliss and Stern comment, then, it is clear that there is a trade-off between a flexible representation of production and the implied rigidity of wages. Nevertheless, as Solow points out, it would be enough of a contribution to macroeconomic theory to demonstrate that wages exhibit some degree of stickiness; complete rigidity of wages is not essential. Thus one could argue that (11) is an approximation of reality, or perhaps that (11) applies in some industries while alternative specifications, such as the more general form in (12), are relevant in other industries. This would enhance

the realism of the production model and at the same time probably preserve some inflexibility in wages.

A frequently observed characteristic of labour markets is a negative correlation between skill levels and unemployment.[21] That is, workers with lower skill levels have higher unemployment rates, other things being equal. According to the dual labour market shirking model, unemployment exists as a disciplinary device in the primary sector because supervision is more difficult in primary sector jobs. Furthermore, jobs in the primary sector are for highly skilled workers. Thus the shirking model apparently predicts a positive relationship between skill level and the unemployment rate, contrary to observation. Akerlof and Yellen (1986) concede that a positive correlation between skill and unemployment is indeed a prediction of the shirking model if all workers are assumed to have the same tastes for work and leisure (shirking). They also argue that there is evidence that the taste for leisure declines with skill level. Akerlof and Yellen propose a slightly modified version of the shirking model in which there are two types of workers: high-skill workers with no taste for shirking, and low-skill workers who do get utility from shirking, as in the original model.[22] Not surprisingly, the outcome of the new assumption is unemployment for low-skill workers, which serves as a discipline device. They conclude that there is a version of the efficiency wage model which is able to explain the observed negative correlation between skill levels and unemployment rates.

Okun's Law states that higher rates of unemployment correspond to lower levels of productivity or, equivalently, that productivity is procyclical. In the Shapiro and Stiglitz (1984) shirking model, a decline in the demand for labour causes equilibrium unemployment to increase. The rise in unemployment increases the cost to workers of losing their jobs, thereby reducing the incentive

[21]The argument below applies to education levels as well to as skill levels.

[22]The reasoning of Akerlof and Yellen is not based on a formal model; they provide an intuitive argument.

to shirk and increasing productivity. Thus it would appear that the simple shirking model predicts countercyclical productivity, contradicting Okun's Law. Once again, Akerlof and Yellen (1986) defend efficiency wage theory in the presence of a prediction seemingly at odds with the facts by proposing an alternative theoretical model. Their analysis is based on a model of long-term (implicit) contracts ignoring efficiency wages. Under certain circumstances the model predicts a positive relationship between output and productivity. Yet the important task of incorporating the long-term contract model into a model of efficiency wages is not undertaken by Akerlof and Yellen.

6 Empirical Tests

Empirical testing of the relatively new theoretical efficiency wage literature has only recently begun and the literature may be divided into two components. One component examines direct evidence for the existence of a relationship between wages and productivity. Most of the evidence is anecdotal in nature. The second component uses multivariate regression techniques to examine evidence for wage differentials across industries, firms or occupations. Recall from the discussion in section 4.1 that evidence of, say, a variation in wages across industries—after controlling for labour quality and compensating variations—is taken as evidence in favour of efficiency wage payments. The direct evidence for a wage-productivity link is discussed in section 6.1 and the evidence for wage differentials is presented in section 6.2.

6.1 Evidence for a wage-productivity link

The essence of the efficiency wage hypothesis is a positive relationship between the level of wages and the productivity of labour. Thus some empirical researchers have set out to examine evidence for a direct link between wages and productivity. Weiss (1980) recounts an event at the Stanford Linear

Accelerator Center which took place in 1975. Management had decided to reduce the work force at the Center by 10 per cent. Workers responded by offering to take a 10 per cent cut in pay in order that layoffs could be avoided. Management, however, rejected the workers' offer on the grounds that pay cuts would cause the best workers to leave. Weiss interprets the management's argument as direct evidence in favour of the adverse selection model of efficiency wages. Akerlof (1982) construes a phenomenon observed by sociologist George Homans in the mid-1950s as evidence supporting the gift-exchange version of efficiency wage theory. Homans studied a group of young female workers doing a job called "cash posting" and noted that the women exceeded the minimum working standards of the firm by an average of 15 per cent. This despite the nature of the job which required little skill and had no prospects for promotion. Akerlof asserts that the only way to rationalize such behaviour is in terms of the gift-exchange model where group work norms are key determinants of worker effort.

Raff and Summers (1987) is a case study of employee compensation, output and profit at the Ford Motor Company of Detroit during the period immediately before and after the introduction of the "five dollar day" in January 1914. Using a variety of sources including contemporary newspaper reports, biographies of Henry Ford, and studies of the mass-production techniques pioneered by Ford, Raff and Summers present a convincing argument that the five dollar day had all the characteristics of an efficiency wage. Prior to 1914, labour turnover and absenteeism at Ford had reached epidemic proportions,[23] and a report commissioned by the Ford management in the summer of 1913 concluded that there was a serious morale problem among production-line workers. In January 1914, the working day at the Ford plant was decreased from 9 to 8 hours and minimum daily pay was increased from $2.34 to $5.00. There is substantial evidence that the new

[23]In 1913, annual turnover at the Ford plant reached 370 per cent and the daily rate of absenteeism was 10 per cent [Raff and Summers, pages S63–S64].

wage was in excess of prevailing market clearing levels; wages at other automobile manufacturers in Detroit were between $2.00 and $3.00 per day, and there were massive queues for jobs at the Ford factory gate. Following the introduction of the five dollar day, turnover and absenteeism rates declined significantly,[24] productivity is estimated to have increased 30–70 per cent and profits continued to increase. Raff and Summers conclude that there is overwhelming evidence that Ford paid what amounts to an efficiency wage in order to reduce turnover and improve worker morale, and that productivity and profits responded in a manner consistent with the efficiency wage hypothesis.

Two criticisms apply to the work of Weiss (1980) and Raff and Summers (1987). First, the evidence presented in these papers is purely circumstantial, and, as such, is a long way from establishing statistically a line of causality from wages to labour productivity. Second, while the work of these authors is testimony to the existence of a positive wage-productivity link in two isolated cases, it is far from widespread evidence of such a link. To support the claim that efficiency wages play a role in macroeconomic fluctuations, one would need evidence of a wage-productivity link in many different sectors of the economy.

Leonard (1987) is an attempt to address the first of the two above criticisms and to examine econometric evidence for the existence of efficiency wages in a particular industry. The Leonard data are taken from a survey of employment conditions for one industry in one U.S. state covering 70,000 workers in 290 occupations.[25] The data include information on the wage level, the level of supervision for selected occupations and the turnover rate for each firm. The shirking model predicts a negative relationship between the level of supervision and the wage rate. Leonard, however, finds no signif-

[24]The turnover rate was 54 per cent in 1914 and 16 per cent in 1915 [Raff and Summers, page S79].

[25]Leonard does not reveal the industry or location from which the data are derived.

icant correlation between the average level of supervision, as proxied by the ratio of supervisors to employees for each occupation, and the average wage. In addition, a regression of wages on supervision level yields no significant relationship between the two variables. The turnover model predicts a negative correlation between wages and turnover rates. Again, Leonard finds no evidence to support the prediction. He does report a substantial variation in wage levels across firms within specific occupations, but concludes that, while this is not inconsistent with the shirking or the turnover models, the variation is not due to either model in the data he uses.

The analysis in Leonard is too simplistic to be regarded as anything more than *prima facie* evidence against the shirking or turnover models. For example, a simple regression with a single independent variable, as in the case of the Leonard model of wages where the level of supervision is the only right-hand side variable, is subject to severe omitted variable problems. Leonard is clearly limited in the scope of his study by the availability of data. Furthermore, as noted above, a more complete test of efficiency wage models is compelled to use data from more than one industry, preferably over a period of time.

6.2 Evidence for interindustry wage differentials

According to competitive theory, wages for identical jobs must be equalized across industries once compensating differentials and quality differences are taken into account. Wage differentials, i.e. wages in excess of the going rate in some industries, can be explained in one of two mutually exclusive ways: (i) firms paying excess wages are behaving irrationally and (ii) firms are paying excess wages because profits increase when wages increase. The second explanation, of course, is a defining feature of efficiency wage models. Thus a commonly used methodology in the empirical efficiency wage literature is to estimate a standard cross-sectional wage equation, control for human capital,

demographic background and working conditions, and then analyze the effect of industry dummy variables on relative wages. Statistically significant industry effects are then interpreted as evidence consistent with efficiency wage theory.

Wage differentials have long been observed by empirical labour economists. Grouping the work of several authors, Leonard (1987) reports that there is evidence of persistent and unexplained wage differentials in the U.S. for almost a century. An oft-cited study is that of Dunlop (1957) which examines the wages of unionized truck drivers in Boston for the year 1951. Dunlop finds that the highest-paid truck driver earns almost twice as much as the lowest-paid driver. Recently there have been several comprehensive studies undertaken in the U.S. of wage structure across industries and occupations. Katz (1986) examines cross-sectional wage data taken from the 1986 Current Population Survey (CPS) for full-time, non-agricultural, private-sector workers. Controlling for education (years of schooling), experience (years in the labour force), occupation, and demographic and location characteristics, Katz finds that workers in the highest-paying industries, mining and transportation, earn 45 per cent and 32 per cent more, respectively, than workers in the lowest-paying industry, retail trade. Using the same data, Dickens and Katz (1987) find that these differences persist when union and non-union workers are analyzed separately, and that the pattern of industry wage premiums is similar for union and non-union workers. Thus it does not appear that the interindustry wage differentials have anything to do with union status of the workers.

A criticism of the above studies is that they have not allowed for unobserved differences in labour quality which could account for the measured wage differentials. The wage differentials may be generated, for example, by differences in technology across industries, which make it profitable to hire higher quality workers and thus pay higher wages in some industries. To the extent that these quality differences are not observed and not ade-

quately controlled for by observed quality variables like education levels, the degree of wage differences between industries may be overestimated. One way to control for unmeasured (time-invariant) labour quality is to use first-difference estimation with longitudinal data. If industries with high wages simply have workers with high levels of unobserved ability, and if ability is rewarded equally across industries, then the wages of workers switching industries should not be systematically linked to observed industry wage differentials. Longitudinal data, of course, permit observation of workers changing industries. And if ability is equally rewarded in all industries, first-difference estimation removes the effect of time-invariant unobserved quality components on estimated wage differences. Murphy and Topel (1987) examine longitudinal data from the CPS covering the period 1977–84. Using industry wage differentials observed in cross-sectional data as the measure of expected wage gains from switching industries and using first-difference estimation, Murphy and Topel find that only 30 per cent of the expected wage gains are realized by workers moving from one industry to another. According to their results, industry wage differentials are primarily due to unobserved differences in labour quality.

Blackburn and Neumark (1988) also examine whether unobserved worker quality differences are responsible for interindustry wage differentials. They point out that attempts to control for ability using first-difference estimation are subject to sample selectivity bias if only the high-ability workers change industries. The alternative approach suggested by Blackburn and Neumark is to incorporate explicit measures of worker ability directly into the estimation of the wage equations. Both IQ test scores and a variable reflecting general knowledge of the labour market are used as measures of ability. Additional data are from the National Longitudinal Survey for non-black males aged 14-24 years in 1966 surveyed in 1973 and again in 1980. Cross-sectional wage equations are estimated for both time periods, with and without the ability variables, allowing for industry- and occupation-specific effects. In

contrast to Murphy and Topel, the results of Blackburn and Neumark indicate that neither interindustry nor interoccupational wage differentials are due to variation in labour quality across industry or occupation cells. Thus the results of Blackburn and Neumark add credence to the findings of Katz (1986) and Dickens and Katz (1987). Clearly, then, there is no consensus evidence that estimated wage differentials are due to pure industry effects or due to the effect of unobserved worker quality.

The most comprehensive study of wage differentials is contained in Kreuger and Summers (1988). The data used in this study are from the CPS for the years 1974, 1979 and 1984, covering full- and part-time non-agricultural workers aged 16 years and over. In the study, the proxy for labour quality is educational attainment; the proxies for compensating differentials are a quantitative measure of working conditions derived from the Quality of Employment Survey: Union and non-union subsamples of the data are examined separately, and transitory shifts in labour demand are accounted for by examining the time periods individually. Given these control variables, Kreuger and Summers still find evidence of substantial interindustry wage differences. In addition, the results indicate that wage differentials are roughly the same for young and old workers, that the differentials are greater in larger firms, and that the differentials are the same across occupations.

Examining wage differentials for the same occupation in different industries is another way of testing efficiency wage theory. Dickens and Katz (1987) use 1983 CPS data for 12 occupations in the non-union private sector to calculate correlations between average wages for workers in any two occupations within an industry. They find that the correlations are between 0.7 and 1.0, similar to the findings of Kreuger and Summers. In other words, if one occupation in an industry is highly paid, all occupations in the industry tend to be highly paid. These results are consistent with shirking or turnover costs differing from one industry to the next. Contrary to Dickens and Katz, Leonard (1987) finds that wage correlations across occupations are quite low

and that some are negative. These results are consistent with shirking or turnover costs differing from one occupation to the next. As for the unobserved worker quality debate, then, there is no consensus on the occupational structure of wage premiums.

Beaudry (1989) dispenses with examining simple wage differentials across industries or occupations and examines directly the empirical evidence in support of his particular version of the turnover model. The model predicts (i) that individual wage profiles for a particular job are independent of the change in wages observed when new employees first start working at the job and (ii) that the initial wage observed when a new employee starts a job is a linear predictor of future wages for the employee at the same job with a coefficient of one. Beaudry uses the National Longitudinal Survey Youth Cohort to construct a panel data set of 401 youths aged 14 to 22 followed for seven years from 1979 to 1985. The empirical specification is designed so that a number of alternative hypotheses about the wage determination process, including a symmetric information version of the original model and a simple market-clearing model, are nested in the estimated wage equation. The data are unable to reject either of the hypotheses implied by the asymmetric information version of the turnover model. Since the specification nests several alternative hypotheses, Beaudry's results are supportive of the theoretical model.

To summarize the empirical work on wage differentials, there are several points worth noting. First, as noted in section 4.1, wage differentials are neither necessary nor sufficient for the existence of efficiency wages. That is, wage differentials may arise in the absence of efficiency wage payments, e.g. due to institutional factors. Alternatively, an absence of wage differentials could be due to similar efficiency wage payments in all firms. Hence, empirical evidence in favour of interindustry or interoccupational wage differentials is not decisive. It is more appropriate to view observed wage differentials as evidence contrary to the competitive model of the labour market rather than

as evidence supporting the efficiency wage hypothesis.

Second, to the extent that interindustry or interoccupational wage differentials do exist, efficiency wage theory makes unambiguous predictions about their origins. The shirking model predicts that wage differentials are due to differences in supervision costs, and the turnover model predicts that differentials arise from differences in turnover costs. A natural line of investigation would be to determine the extent to which wage differentials are explained by differences in supervision and turnover costs across firms. A study along this line would be able to determine whether higher wages are associated with higher supervision and turnover costs, and the degree of correlation between wages and supervision and turnover costs across industries. Such a study would prove useful in determining which of the efficiency wage models is more empirically relevant. The recent study by Beaudry (1989) is an important step in this area.

Finally, regardless of the origins of efficiency wages, the efficiency wage hypothesis predicts a positive relationship between labour productivity and the wage rate. An alternative empirical approach would be to investigate a large-scale micro data set for evidence of the wage-productivity link. An example of this type of approach, in a different context, is the work of Brown and Medoff (1978) which estimates a production function in order to examine the effect of labour unions on productivity.

7 Conclusion

In the mid-1970s, implicit contract theory [Baily (1974), Azariadis (1975)] was proposed as an explanation of rigid real wages and underemployment equilibriums. After more than a decade of research, it has become apparent that implicit contract theory predicts overemployment more easily than underemployment, and that the types of contract predicted by the theory are

simply not observed.[26] Whether efficiency wage theory will be accorded the same fate, of course, remains to be seen. Proponents view efficiency wage models as capable of accounting for downwardly rigid real wages, involuntary unemployment, wage differentials for identical workers and dual labour markets. Opponents consider the bonding critique sufficiently fundamental to undermine the theoretical validity of efficiency wage theory.

Akerlof and Yellen (1986), Stiglitz (1986), and Katz (1986) all agree that existing models of efficiency wages require more explicit consideration of the long-term relationship between employer and employee. Incorporating efficiency wages into a model of long-term contracts is potentially fruitful for two reasons. First, the interaction between efficiency wages and implicit bonding should become more clear. In this way it may become apparent under which circumstances efficiency wage payments will arise and under which circumstances bonding schemes will arise. Second, an explicit dynamic framework will be useful for examining the role of efficiency wages in cyclical fluctuations.[27]

Testing the empirical predictions of efficiency wage theory is still at an early stage. Most of the existing evidence supporting a direct and positive relationship between wages and productivity is based on isolated case studies, or on anecdote. The remainder of the empirical work focuses on the existence of wage differentials for similar workers, despite the fact that wage differentials are neither necessary nor sufficient for the existence of efficiency wages. Given the current theoretical difficulties faced by efficiency wage models, it seems likely that the most important work on the topic in the near future will be in the area of empirical research. Ideally, this research will identify which of the theoretical models of efficiency wages is most applicable, which of the predictions of the theory are consistent with the observed facts and

[26] For a recent critical appraisal of implicit contract theory, see Stiglitz (1986).

[27] Kimball (1989) derives the labour market dynamics of the Shapiro and Stiglitz shirking model.

whether there is indeed a relationship between worker productivity and the wage level.

References

Akerlof, G.A., 1982. "Labor Contracts as a Partial Gift Exchange," *Quarterly Journal of Economics*, **97**: 543–569.

Akerlof, G.A. and J.L. Yellen, 1985. "A Near Rational Model of the Business Cycle with Wage and Price Inertia," *Quarterly Journal of Economics*, **100**: 823–838.

_____ , eds., 1986. *Efficiency Wage Models of the Labor Market*, New York: Cambridge University Press.

Azariadis, C., 1975. "Implicit Contracts and Underemployment Equilibria," *Journal of Political Economy*, **83**: 1183–1202.

Baily, M.N., 1974. "Wages and Employment Under Certain Demands," *Review of Economic Studies*, **41**: 37–50.

Beaudry, P., 1989. "Entry Wages Signalling Future Wages: Theory and Evidence on the Relationship between Firm-Specific Capital, Wage Profiles and Job Rationing," Cahier 8905, Départment de sciences économiques, Université de Montréal.

Blackburn, M. and D. Neumark, 1988. "Efficiency Wages, Inter-Industry Wage Differentials and the Returns to Ability," Finance and Economics Discussion Series No.32, U.S. Board of Governors of the Federal Reserve System.

Bliss, C. and N. Stern, 1978a. "Productivity, Wages and Nutrition. Part I: The Theory," *Journal of Development Economics*, **5**: 331–362.

_____ , 1978b. "Productivity, Wages and Nutrition. Part II: Some Observations," *Journal of Development Economics*, **5**: 363–398.

Bowles, S., 1985. "The Production Process in a Competitive Economy: Walrasian, Neo-Hobbesian, and Marxian Models," *American Economic Review*, **75**: 16–36.

Brown, C. and J. Medoff, 1978. "Trade Unions in the Production Process," *Journal of Political Economy*, **86**: 355–378.

Bulow, J.I. and L.H. Summers, 1986. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination, and Keynesian Unemployment," *Journal of Labor Economics*, **4**: 376–414.

Calvo, G., 1979. "Quasi-Walrasian Theories of Unemployment," *American Economic Review*, **69**: 102–107.

Carmicheal, L., 1983. "Firm-Specific Human Capital and Promotion Ladders," *Bell Journal of Economics*, **14**: 251–258.

_____ , 1985. "Can Unemployment Be Involuntary?: Comment," *American Economic Review*, **75**: 1213–1214.

_____ , 1987. "Efficiency Wage Models of Unemployment: A Survey," Unpublished manuscript, Queen's University.

Diamond, P.A., 1981. "Mobility Costs, Frictional Unemployment, and Efficiency," *Journal of Political Economy*, *89*: 798–812.

_____ , 1982. "Aggregate Demand Management in Search Equilibrium," *Journal of Political Economy*, **90**: 881–894.

Dickens, W.T. and L.F. Katz, 1987. "Inter-Industry Wage Differences and Industry Characteristics," in *Unemployment and the Structure of Labor Markets*, K. Lang and J.S. Leonard, eds., New York: Basil Blackwell, Inc.

Dickens, W.T. and K. Lang, 1985a. "A Test of Dual Labor Market Theory," *American Economic Review*, **75**: 792–805.

_____ , 1985b. "Testing Dual Labor Market Theory: A Reconsideration of the Evidence," NBER Working Paper No. 1670.

Doeringer, P.B. and M.J. Piore, 1971. *Internal Labor Markets and Manpower Analysis*, Lexington, Ma.: D.C. Heath.

Dunlop, J.T., 1957. "The Task of Contemporary Wage Theory," in *The Theory of Wage Determination*, J.T. Dunlop, ed., New York: St. Martin's Press.

Eaton, B.C., and W. White, 1982. "Agent Compensation and the Limits of Bonding," *Economic Inquiry*, **20**: 330–343.

Foster, J. and H. Wan, 1984. "Involuntary Unemployment as a Principal-Agent Equilibrium," *American Economic Review*, **74**: 476–484.

Gibbons, R. and L.F. Katz, 1989. "Layoffs and Lemons," NBER Working Paper No. 2968.

Hicks, J., 1932. *The Theory of Wages*, London: Macmillan and Co. Ltd.

Howitt, P., 1985. "Transaction Costs in the Theory of Unemployment," *American Economic Review*, **75**: 88–100.

Jones, S.R.G., 1985. "Dual Labour Markets, Productivity and Unemployment," Discussion Paper No. 85–22, Department of Economics, The University of British Columbia.

Katz, L.F., 1986. "Efficiency Wage Theories: A Partial Evaluation," in *NBER Macroeconomics Manual 1986*, S. Fischer, ed., Cambridge, Ma.: MIT Press.

Kimball, M. S., 1989. "Labor Market Dynamics when Unemployment is a Worker Discipline Device," NBER Working Paper No. 2967.

Krueger, A.B. and L.H. Summers, 1988. "Efficiency Wages and the Inter-Industry Wage Structure," *Econometrica*, **56**: 259–293.

Lang, K. and J.S. Leonard, eds., 1987. *Unemployment and the Structure of Labor Markets*, New York: Basil Blackwell Inc.

Lazear, E.P., 1981. "Agency, Earnings Profiles, Productivity and Hours Restrictions," *American Economic Review*, **71**: 606–620.

Leibenstein, H., 1957. *Economic Backwardness and Economic Growth*, New York: John Wiley and Sons Inc.

Leonard, J.S., 1987. "Carrots and Sticks: Pay, Supervision and Turnover," *Journal of Labor Economics*, **5**: S136–S152.

Malcomson, J., 1984. "Work Incentives, Hierarchy and Internal Labor Markets," *Journal of Political Economy*, **92**: 486–507.

Mirrlees, J.A., 1976. "A Pure Theory of Under-Developed Economies," in *Agriculture in Development Theory*, L. Reynolds, ed., New Haven, Ct.: Yale University Press.

Miyazaki, H., 1984. "Work Norms and Involuntary Unemployment," *Quarterly Journal of Economics*, **99**: 297–311.

Murphy, K.M., and R.H. Topel, 1987. "Unemployment, Risk and Earnings: Testing for Equalizing Differences in the Labor Market," in *Unemployment and the Structure of Labor Markets*, K. Lang and J.S. Leonard, eds., New York: Basil Blackwell, Inc.

Phelps, E.S. et al, 1970. *Microeconomic Foundations of Keynesian Macroeconomics*, New York: Norton Inc.

Raff, D.M.G. and L.H. Summers, 1987. "Did Henry Ford Pay Efficiency Wages?" *Journal of Labor Economics*, **5**: S57–S86.

Salop, S.C., 1979. "A Model of the Natural Rate of Unemployment," *American Economic Review*, **69**: 117–125.

Shapiro, C. and J.E. Stiglitz, 1984. "Equilibrium Unemployment as a Worker Discipline Device," *American Economic Review*, **74**: 433–444.

_____ , 1985. "Can Unemployment Be Involuntary?: Reply," *American Economic Review*, **75**: 1215–1217.

Solow, R.M., 1979. "Another Possible Source of Wage Stickiness," *Journal of Macroeconomics*, **1**: 79–82.

Stiglitz, J.E., 1974. "Wage Determination and Unemployment in L.D.C.'s: The Labor Turnover Model," *Quarterly Journal of Economics*, **88**: 194–227.

_____ , 1976a. "The Efficiency Wage Hypothesis, Surplus Labour and the Distribution of Income in L.D.C.'s," *Oxford Economic Papers*, **28**: 185–207.

_____ , 1976b. "Prices and Queues as Screening Devices in Competitive Markets," IMSSS Technical Report No. 212, Stanford University.

_____ , 1986. "Theories of Wage Rigidity," in *Keynes' Economic Legacy*, J.L. Butkiewicz, K.J. Koford and J.B. Miller, eds., New York: Praeger Publishers.

Summers, L.H., 1988. "Relative Wages, Efficiency Wages, and Keynesian Un-

employment," *American Economics Association Papers and Proceedings*, **78**: 383–388.

Vahey, S.P., 1987. "Efficiency Wage Models: A Literature Survey," Unpublished M.A. Dissertation, Essex University.

Varian, H.R., 1978. *Microeconomic Analysis*, New York: W.W. Norton & Company, Inc.

Weiss, A., 1980. "Job Queues and Layoffs in Labor Markets with Flexible Wages," *Journal of Political Economy*, **88**: 526–538.

Weiss, L., 1986. "Comment on 'Efficiency Wages—A Partial Evaluation' by L.F. Katz," in *NBER Macroeconomics Manual 1986*, S. Fischer, ed., Cambridge, Ma.: MIT Press.

Weisskopf, T., Bowles, S., and D. Gordon, 1983. "Hearts and Minds: A Social Model of Aggregate Productivity Growth in the U.S., 1948–1979," *Brookings Papers on Economic Activity*, **2**: 381–441.

Yellen, J.L., 1984. "Efficiency Wage Models of Unemployment," *American Economics Association Papers and Proceedings*, **74**: 200–205.

Bank of Canada Working Papers

	Title	Author
88-1	On Conditional Rules for Monetary Policy in a Small Open Economy	S. Poloz
88-2	Demande de fonds mutuels, co-intégration et modèle de correction des erreurs	J.-F. Fillion
89-1	A Regime Switching Approach to Uncovered Interest Parity	S. van Norden
89-2	Monthly Estimates of Canadian GDP and its Price Deflator	S. Kozicki
89-3	Estimates of the NAIRU Using an Extended Okun's Law	R. Ford and D. Rose
89-4	On the Aggregate Implications of Optimal Price Adjustment	B. Cozier
89-5	Efficiency Wages: A Literature Survey	T. Fisher

Single copies of Bank of Canada papers may be obtained from:

Publications Distribution
Bank of Canada
234 Wellington Street
Ottawa, Ontario
K1A 0G9