NBER WORKING PAPER SERIES

ARE EFFICIENCY WAGES EFFICIENT?

William T. Dickens

Lawrence Katz

Kevin Lang

Working Paper No. 1935

Are Efficiency Wages Efficient?

ABSTRACT

Efficiency wage models have been criticized because worker malfeasance can be prevented in a pareto efficient manner by requiring workers to post a bond which they lose if they are caught cheating. However, since it is costly to monitor workers and costless to demand a larger bond, firms should pay nothing for monitoring and demand very large bonds. Since we observe that firms devote considerable resources to monitoring workers, bonds must be limited. Therefore firms must use second best alternatives -- intensive monitoring and/or efficiency wages. The payment of efficiency wages cannot be ruled out on a priori theoretical grounds.

William T. Dickens
Sloan School of Management
M.I.T.
Cambridge, MA 0213B

Lawrence Katz
School of Business
 Administration
University of California
Berkeley, CA 94720

Kevin Lang
School of Social
 Science
University of California
Irvine, CA 92717

ARE EFFICIENCY WAGES EFFICIENT?

1/31/86

Efficiency wage models, in which workers are paid a wage above the reservation wage in order to give them an incentive not to quit, steal or shirk, provide an attractive explanation for a wide variety of phenomena including involuntary un- employment, the large differences between industries in the wages paid for what appear to be similar workers, and the existence of the dual labor market. The primary criticism of efficiency wage models is that similar incentives can be provided in a pareto efficient manner by having workers post bonds against malfeasance. Equivalently, firms can pay efficiency wages but charge workers an employment fee which makes them indifferent between taking a particular job or seeking alternative employment. The bond or employment fee can either be paid at the beginning of employment, or over time if workers receive lower wages in the initial months or years of employment. Bonds can be paid back as a lump-sum upon retirement, as an annuity or pension after retirement and/or as higher wages paid most commonly towards the end of one's career. Lazear (1979,1981) has pro- posed this as an explanation for tenure-earnings profiles and a number of other phenomena.

In this note we argue that the first best efficient bonding solution is ev- idently unattainable in practice. With efficient bonding firms should be spending nothing or only an infinitesimal amount on monitoring workers to detect malfeasance. Since firms are observed to devote substantial resources to moni- toring workers, bonding must be incomplete. We argue that firms must be choosing one or both of two inefficient solutions -- monitoring workers intensively and/or paying efficiency wages.[1] Evidently there are limits on the size of the bond a firm may demand. These could be due to either the possibility of the firm de-

faulting on the employment contract or wealth and credit market constraints on workers.

The next section contains a more detailed discussion of this basic point. Section II presents a formal model of bonding and efficiency wages with variable monitoring costs. The formal model allows us to identify the conditions under which firms will choose intensive monitoring rather than efficiency wages.

I. EFFICIENCY WAGES VS. BONDING

Since in a perfectly competitive labor market there is no unemployment and all equivalent workers receive the same wage, separations are costless to workers. As a consequence firms cannot punish worker malfeasance by firing them, and therefore workers can be expected to shirk or steal unless firms can develop other means of punishing malfeasance. Several authors (Calvo, 1979; Stoft, 1982; Shapiro and Stiglitz, 1984; Bowles, 1985; Jones, 1985; Bulow and Summers, 1985) have developed models in which firms pay workers more than their reservation wages so that it is costly to workers to be fired (or equivalently, in models in which workers quit too frequently, it is costly to them to quit).

A common objection to these models is that employers can solve these problems in a pareto-efficient manner by requiring employees to post a bond against shirking or quitting. Instances of direct bonding are very rare although not unknown. However, overt bonding is formally equivalent to the sort of employment relation described by Lazear (1979, 1981). In his model, workers post a bond against cheating by accepting wages below their marginal product initially and

have it returned in the form of wages above the value of their marginal product later in life and/or in the form of a pension upon retirement. Many find this argument -- that efficiency wages are inefficient and therefore unlikely to be used -- compelling.[2] Discussion at several conferences and seminars we have attended indicates to us that acceptance of the efficiency wage model has been inhibited by this argument.[3]

A common response to this objection is to argue that there are limits on the size of the bonds firms can demand. An alternative to paying efficiency wages when bonding is limited is increased monitoring of workers. Some monitoring is required with either bonding or efficiency wages since it is useless to make it costly to workers to be caught cheating if they are never caught. Monitoring by itself is insufficient to deter malfeasance since there must be some cost to workers of being caught. Considering only the size of the bond and monitoring expenditures, the firm faces an optimal law enforcement problem in which it must choose the amount of resources to devote to catching malfeasants and the punishment to inflict on those it catches. The solution to this problem is well known (Carr-Hill and Stern, 1979; Stern, 1978) and has been aptly characterized as "hang litterers with probability zero." That is, next to nothing should be spent on catching law breakers since it is cheaper to deter crime by making punishments more severe. This solution applies equally to the bonding model since it is costly to the firm to monitor workers but costless to increase the bond. As long as any resources are being devoted to monitoring, the same degree of deterrence can always be attained at lower cost by increasing the bond and reducing the level of monitoring. Thus the efficient bonding solution involves no expenditure on monitoring. If the probability of detection is near zero when the firm does not monitor workers, the bond required to prevent malfeasance will be very large.[4]

There are two strong logical reasons for believing that the feasible bond may not be as large as necessary if firms spend little or nothing on monitoring. The first is that, particularly at the beginning of their work lives, workers face capital market constraints and thus may not be able to afford large bonds. The second is that if the bond were very large, the firm would have a strong incentive to claim malfeasance on the part of the worker and keep the bond. There is also empirical evidence that suggests one or the other of these two constraints is commonly binding. Since the optimal expenditure on monitoring is zero, if we observe any expenditure on monitoring by firms, the efficient bonding solution must not be obtainable, and it must be the case that the optimal bond exceeds the maximum imposed by either capital markets or incentives for firm malfeasance. Cursory inspection of almost any workplace indicates that firms do expend substantial resources monitoring workers. Firms hire supervisors, auditors, inspectors and testers. The pace of many jobs is monitored by machines. Evidently firms must choose second-best methods of avoiding worker malfeasance -- intensive monitoring and/or efficiency wages. Which technique firms choose will depend on the monitoring technology.

II. A FORMAL MODEL

We use a one period model. At the beginning of the period workers may post a bond, b. During the period workers may or may not supply effort, e, which takes on the value 1 if effort is supplied and 0 otherwise. The probability, $p(m)$, that any shirking which occurs is detected by the firm depends on the firm's expendi-

ture on monitoring, m. We assume $p'>0$ and $p(0)=p_{min}>0$. If the workers are caught shirking, they forfeit the bond and do not receive the wage, w. However, they are immediately employed elsewhere at the reservation wage, r. If they are not caught shirking, they receive the wage and the bond is returned.

It should be obvious that the only possible equilibria are those in which e equals 1. If workers provided zero effort, no output would be produced. The firm would find it profitable to hire workers only if the expected wage net of foregone bonds were nonpositive. However, with a nonpositive wage, workers will be unwilling to supply labor.

Therefore, the firm's objective is to maximize profits, π, which are given by

$$(1) \qquad \pi = R(L) - (w+m)L$$

where L is the number of workers employed by the firm and R is revenue net of nonlabor costs ($R'>0$, $R''<0$). Profits are maximized subject to two constraints. The first is that the wage exceeds the reservation wage:

$$(2) \qquad w \geq r.$$

The second is that workers must not have an incentive to shirk.

In specifying the conditions under which a worker will shirk, we assume that workers are risk neutral. The model could easily be adjusted to allow for risk aversion. We also assume, as noted above that workers who are fired for shirking can earn the reservation wage for the remainder of the period. This is the equivalent of the standard competitive assumption that workers who are fired immediately obtain employment at the competitive wage rate. Workers will not shirk

if the wage plus bond they receive if they do not shirk exceeds the expected wage plus bond if they do shirk by at least the value of shirking. Thus we have

(3) $w + b \geq v + (1-p(m))(w+b) + p(m)r$

where v is the benefit from shirking. There is no benefit to combinations of w and b which exceed their minimum level so we treat (3) as an equality. Rearranging terms gives

(4) $w + b = v/p(m) + r.$

        The firm chooses L, w, m and b to maximize (1) subject to inequality constraint (2) and the equality constraint (4). In such problems, the constraint that the choice variables be nonnegative can usually be ignored and an interior solution assumed. In this case, however, the requirement that m be nonnegative may be a binding constraint. We therefore include this constraint explicitly. The first order conditions are given by

(5) $\partial(.)/\partial L = R' - m - w = 0$

(6) $\partial(.)/\partial w = - L + \mu + \lambda = 0$

(7) $\partial(.)/\partial m = - L + \lambda vp'/p^2 + \psi = 0$

(8) $\partial(.)/\partial b = \lambda = 0$

where ψ is the Kuhn-Tucker multiplier associated with the constraint that m be nonnegative, λ is the Lagrange multiplier associated with (4) and μ is the Kuhn-Tucker multiplier associated with (2).

        The solution to (5) - (8) is

(9) $m = 0$

(10) $R' = w = r$

(11) $b = v/p(0)$.

Thus the efficient solution involves bonding with no monitoring. Efficiency is achieved in competitive markets because the number of workers hired is chosen so as to set the marginal revenue product equal to the wage.

As noted in the last section, firms do spend resources on monitoring workers; so there must be limits on the extent to which bonding can be used. A formal model of the processes which limit the size of bonds is beyond the scope of this note.[5] For the purposes at hand it suffices to assume that b cannot exceed some maximum value, b*. In this case, the firm maximizes profits subject to inequality (2), equality (4), and $b = b*$. Setting $b = b*$ and reexamining (5) - (7), it is evident that there are three possible solutions corresponding to the cases where (6) holds with equality, (7) holds with equality and both (6) and (7) hold with equality. Thus the firm may choose to use only increased monitoring, only an efficiency wage or both. In the case where both are used, the level of monitoring is chosen to set vp'/p^2 equal to one. In sum the relevance of efficiency wages is an empirical not a theoretical question.

III. CONCLUSIONS

The critical weakness in the argument that efficiency wages are an inefficient solution to the problem of preventing worker malfeasance and are therefore unlikely to be used in practice is that the size of bonds that workers can post is limited. The fact that firms expend significant resources monitoring workers indicates that bonds are limited in practice.

One obvious factor restricting the size of the bonds workers can post is wealth and credit market constraints. Another potential problem is incentive compatibility problems on the firm's side. Lazear (1979, 1981) recognizes that if a worker's wage exceeds his VMP, the firm will have an incentive to fire the worker. Lazear relies on reputation costs to prevent the firm from responding to this incentive. However, reputation costs may not be very large. The most a firm can sacrifice if it loses its reputation is the present value of future profits and any transaction costs involved in selling off its capital stock. In a competitive market, the present value of expected future profits will be near zero, and thus in a competitive market reputation costs should not exceed the transactions costs for transferring an enterprise to new owners or hiring new management.

Moreover, if economic actors have rational expectations, past actions are only signals of future behavior if firms differ in some relevant way which is unobservable so that defaulting conveys some new information. In models which assume that all firms are ex ante identical or that workers have perfect information about firm characteristics, if reputation costs are possible at all, they rest on the etherial assumption that reputations matter because they matter.[6]

Thus it is not surprising that firms do not rely entirely on bonding to prevent malfeasance but instead devote resources to monitoring workers. Since resources are devoted to monitoring workers, depending on the monitoring technology, it may pay firms to reduce monitoring costs by paying efficiency wages. Efficiency wages are not necessarily inefficient.

FOOTNOTES

1. Eaton and White (1983) analyze the trade-off between efficiency wage payments and increased monitoring effort when no bonding arrangements are possible.

2. We refer here to private efficiency relative to the use of alternative mechanisms for controlling worker malfeasance rather than to the Pareto optimality of the market outcome. The market equilibrium is generally not efficient since firms fail to adequately take into account the impacts of their wage and monitoring levels on the policies other firms must utilize to prevent shirking by employees (Shapiro and Stiglitz; 1984, 1985).

3. It should be noted that the bonding critique does not apply to norm based efficiency wage models (Solow, 1979; Akerlof, 1984), to adverse selection based models in which firms never learn a worker's true ability (Weiss, 1980), or nutrition models (Stiglitz, 1974).

4. The possibility of mistakes (Type II errors) may be another factor limiting the size of the bond a firm can demand. If firms sometimes fire workers who were not cheating, a very high bond may no longer be optimal since the firm would have to compensate workers for the possibility that they would lose their bond due to an error. The larger the bond the larger the necessary compensation. This is only a problem if the monitoring technology doesn't allow the firm to know with complete certainty that a worker was cheating in at least some cases. If the firm can know some cases with certainty, the best thing to do, if workers are at all risk averse, is to fire only those it knows with certainty have shirked and to set the bond as high as needed to avoid shirking given the probability of detection with certainty. If workers are risk neutral, then the best solution will be perfect bonding with workers being compensated for the possibility of a mistake with a wage premium exactly equal to their expected loss. If workers are risk averse and some mistakes are inevitable, the firm may find it optimal to expend resources on monitoring to reduce the required bond and compensating wage differential. When mistakes are unavoidable, some of the attractive features of the efficiency wage model may still obtain even if there are no queues for jobs. Industry wage differentials can be explained as compensation for the accidental loss of a bond. In any case, our basic argument, that if we observe expenditures on monitoring, then perfect bonding is not attainable holds. One further caveat is that firms may monitor workers for other reasons than perventing worker malfeasance. Monitoring may be utilzed to avoid honest worker errors or to help in sorting heterogeneous workers. Still, we do believe that one does observe monitoring activities that plainly serve to help deter shirking, carelesness and stealing by workers.

5. Eaton and White (1982) show how capital market imperfections can limit the use of bonding and give rise to efficiency wage payments. They assume exogenous shirking detection possibilities.

6. The likelihood of firm default can be reduced if the firm does not expect to gain anything from a default. For example, if a firm claims that a worker has shirked and fires the worker it is possible to pay the bond to a third party instead of the firm. Then the worker is disciplined but the firm gains nothing if it cheats. While this is a theoretical possibility we do not observe the use of this device. One arrangement which seems at first to resemble this is having payments made to a pension fund (at least up to the point where the pension is vested). However, if the firm defaults it does benefit. Either it can reduce its future payments to the fund or the fund will pay higher benefits so that the firm can pay a lower wage and still attract workers.

Footnotes

BIBLIOGRAPHY

Akerlof, George. "Gift Exchange and Efficiency Wages: Four Views," American Economic Review Proceedings, 74 (May 1984), 79-83.

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

Bulow, J. I. and Summers, L. H. "A Theory of Dual Labor Markets with Application to Industrial Policy, Discrimination and Keynesian Unemployment," NBER working paper #1666 (July,1985).

Calvo, G. "Quasi-Walrasian Theories of Unemployment," American Economic Review, May 1979, p102-7.

Carmichael, Lorne. "Can Unemployment Be Involuntary?: Comment," American Economic Review, 75 (December 1985), 1213-14.

Carr-Hill, Roy A. and Nicholas H. Stern. Crime, The Police, and Criminal Statistics, London: Academic Press, 1979.

Eaton, C. and W.D. White. "Agent Compensation and the Limits of Bonding," Economic Inquiry, (July 1982), 330-343.

------------------------, "The Economy of High Wages: An Agency Problem," Economica, (April 1983), 175-81.

Jones, Stephen R.G. "Minimum Wage Legislation in a Dual Market," mimeo, University of British Columbia, May 1985.

Lazear, E. "Agency, Earnings Profiles, Productivity, and Hours Restrictions," American Economic Review, (September 1981), 606-620.

--------------, "Why is there Mandatory Retirement?" Journal of Political Economy, 87(6), (December, 1979), 261-84.

Shapiro, C. and J.E. Stiglitz, "Equilibrium Unemployment as a Worker Discipline Device," American Economic Review, 74, (June 1984), 433-44.

----------------------------, "Can Unemployment be Involuntary?: Reply," American Economic Review, 75, (December 1985), 1215-1217.

Solow, R. "Another Possible Source of Wage Stickiness," Journal of Macro Economics, 1, 1979, 79-82.

Stern, Nicholas H. "On the Economic Theory of Policy Towards Crime," in J. M. Heineke (ed.) Economic Models of Criminal Behavior, Amsterdam: North-Holland, 1978.

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

Stoft, S., "Cheat Threat Theory: An Explanation of Involuntary Unemployment," mimeo, Boston University, May 1982.

Weiss, Andrew, "Job Queues and Layoffs in Labor Markets with Flexible Wages," Journal of Political Economy, 88, (June 1980), 526-38.

Yellen, Janet L. "Efficiency Wage Models of Unemployment," American Economic Review Proceedings, 74, (May,1984) 200-5.

Bibliography