Lecture 19,   2205
Efficient Wages and Implicit Contracts   R. Butler

I. Efficiency Wages

In all of the contracts considered last period, firms compensation policies were influenced by market constraints: if the time-rate was too low, firms couldn't attract workers; if the tournament prizes were too high, other firms would enter and compete away the business. Delayed compensation schemes were similarly disciplined in the market. The idea of efficiency wages is to pay wages in excess of the market in order to induce greater productivity of the workers. This is most convincingly argued in payment schemes in less-developed economies. If the worker is just paid a subsistence wage, then they may not be nourished enough to work hard. If they are paid above the subsistence level, the firm may enhance workers productivity and increase its profits if the additional output from the workers offsets the extra wage compensation.

While the less-developed nation version of efficiency wages seems reasonable, there is some controversy whether efficiency wages are paid in the U.S. economy. The controversy centers around the **bonding critique**, which suggests that efficiency wages should self-destruct in the long run. The reason for this is that—unlike the other schemes considered last lecture in which the workers were paid their MRP (in the long run at least)—efficiency wages differ in that firms determine the efficiency wage without regard to market conditions (or without regard to the constraints placed on the firm by the marketplace). So paying an efficiency wage, above the market clearing wage, should result in too many job applicants. These additional job applicants are willing to compete for the wages with the higher wages; one way that they could compete is to post a bond at the time they were hired. If the firm catches them cheating, then it fires them and keeps the bond. If the worker is not found to cheat and maintains his or her employment with the firm, then the firm would return the bond to the worker (with accumulated interest) at the time of the worker's retirement.

These employment bonds are not observed, though we discussed delayed compensation schemes that could serve the same purpose. But, under these schemes, there would be no need to pay wages that were above the competitive wage level over the whole person's lifetime (i.e., the overpayments later would match the underpayments now). Hence, the importance of efficiency wages (and the bonding critique) are still widely debated in the profession.

Now for the efficiency wage model, we start with profit maximizing behavior, where the firm now chooses both the number of workers, N, and a wage rate, w, in order to maximize profits. Wage is endogeneous because worker effort is assumed to rise with wages (effort=E, and E=E(w) such that $\frac{\partial E}{\partial w} > 0$), and output depends both on the number of workers and on their effort. Hence, profit maximization is as follows:

$$\text{profit} = P*f(N*E(w)) - w*N$$

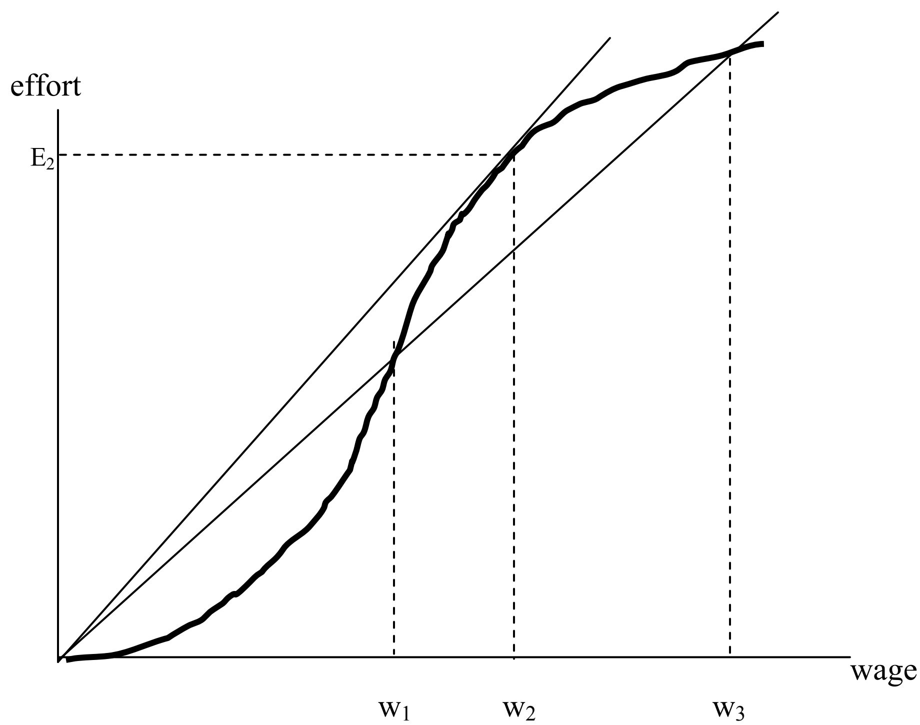
where P=product price, f(.) is the production function, N=number of workers, E=effort made by each worker, and w=wage paid by the firm. The two first order conditions for profit maximization are

P*f'*E(w) = w (the partial derivative with respect to N)
P*f'*N*E' = N (the partial derivative with respect to w)

where f' , E' indicate partial derivatives of the respective functions. Combining these two first order conditions we get the following result:

$$\frac{\partial E}{\partial w}\frac{w}{E}=1$$

This implies that the wage should be set at the point with the elasticity of effort with respect to the wage rate is 1. The intuition behind this result can be seen by examining the relationship between effort and the wage rate as follows:



From the production function given above, the marginal productivity of labor increases proportionally with respect to effort. Effort is assumed to increase with respect to the wage as illustrated—increasing more than proportionately with the wage initially, then eventually showing diminishing returns. At wage w_1, the change in effort with respect to the change in wage (i.e., the marginal change in effort as given by the slope of the thick black, effort-wage function) is greater than the average effort per wage (i.e., the average change in effort as given by the slope of the line from the origin to the point where w_1 intersects the thick black, effort-wage function). $\frac{\partial E}{\partial w}\frac{w}{E}$ is greater than one at this point, so that by increasing the wage beyond w_1, effort will increase proportionality more than the wage will increase. This will increase firm profitability. Similarly, $\frac{\partial E}{\partial w}\frac{w}{E}>1$ for all points between w_1 and w_2 and so it is optimal to increase the wage to w_2. At w_2, $\frac{\partial E}{\partial w}\frac{w}{E}=1$, so that the marginal change in effort just equals the average change in effort. For points beyond w_2, say those points between w_2 and w_3, $\frac{\partial E}{\partial w}\frac{w}{E}$ is less than one, implying that additional increases in wages are accompanied by less than proportional increases in output. Hence, w_2 is the optimal wage, with E_2 the profit maximizing level of output.

The profit maximizing firm sets w_2 independently of market forces, except that the firm cannot set the wage lower than the market wage or it will fail to attract workers. Therefore, those firms employing efficiency wages will have an oversupply of job applicants. However, the firm will not have an incentive to reduce the wage (given our model assumptions); as a reduction in wage will reduce effort (and output) by more than the reduction in wage. Hence, some workers will be unemployed (and this would help to explain both persistent unemployment, and persistent wage differences across industries).

II. Implicit Contracts (from Sherwin Rosen's "Implicit Contracts—A Survey" in *Handbook of Labor Economics*)

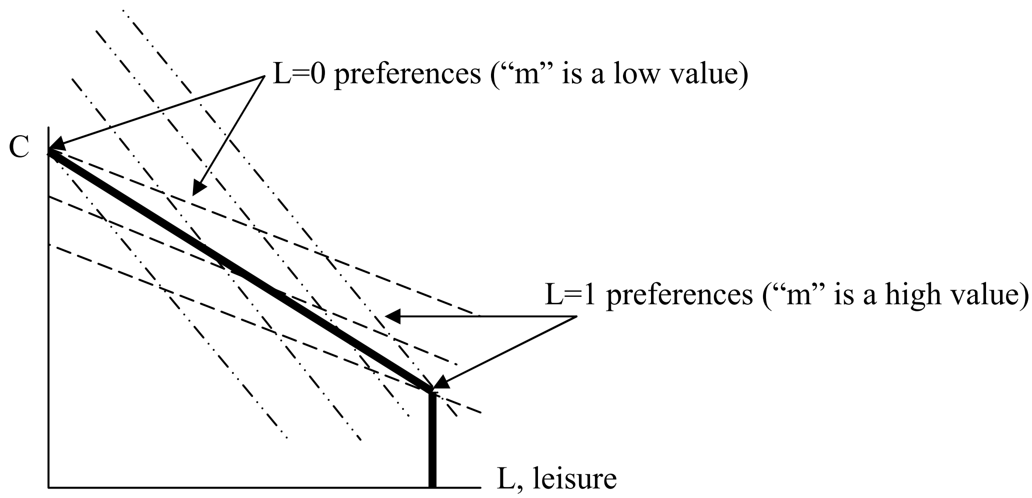
The contract exists because the worker has some specific capital that the firm will find expensive to replace, and because there is some uncertainty about the future--firm demand is subject to some random fluctuation, and so the derived demand for labor is also uncertain. Workers are risk averse (so they are willing to pay to resolve some of the uncertainty about their future), and are willing to enter into an agreement with the firm about the conditions under which they will be laid off.

One way to treat uncertainty is to say that the price that the firm receives for its product, θ, behaves like a random variable in the sense that its value in the next period (and in any future period) is not known with certainty, though we have some idea about what values it might take. (While we don't know its exact value, we know what distribution from which the price is being drawn.) The contract is a set of conditions such that "if θ has a value θ_1, then worker agrees to work ZZZZ hours for which the firm will pay the worker XXXX dollars." Given the various potential values for θ, how does the worker and firm work out its implicit contract?

Rosen demonstrates this procedure with a particularly simple example: the worker has straight line indifference curves, with a constant marginal rate of substitution equal to "m" (Can you prove this for the utility function given below?). That is, the utility function of the worker looks like

U = U(C+mL)

where C is the level of consumption and L is the fraction of time spent in leisure (0<=L<=1). With this utility function, what does the labor supply function look like? Assume that the utility function is concave, so that $\frac{\partial U}{\partial (C+mL)} > 0$ and $\frac{\partial^2 U}{\partial (C+mL)^2} < 0$. Note that C and L are perfect substitutes (show that the indifference curves are straight lines), so that for any given wage, the worker is either working full time or not at all, depending on the slope of her indifference curves:



Recall that the slope of the dark, budget line is –w/p (while the slope of the indifference curves is –m). All workers choose work full time (L=0) if w/p>m, and they choose full time leisure (L=1) if w/p<m. The worker is assumed to have an expected utility function that is written analogously to the expected value of a random variable. The is, if there are 2 outputs for a random variable X (say X_1 and X_2) which occur with probability p and 1-p respectively, than the "expected **value**" of X is

expected value of X = p*X_1 + (1-p)*X_2

Similarly, the "expected **utility**" of a worker facing two uncertain outcomes in the future is just (say p is the probability of not being laid off, so that 1-p is the probability of a lay off)

expected utility = p*U(C_1+mL_1) + (1-p)*U(C_2+mL_2)

where we assume the following:

* there is no nonwage income, so that
* C_1 is the wage payment when someone works
* C_2 is the unemployment payment (by the firm) for those who are laid off
* L_1=0 and L_2=1 (these are not assumptions, but consequences of the assumed form of the utility function. Do you understand why?)

* that utility is only a function of the level of C and L and not whether or not one is working (this simplifies the model, but what does it imply?)

For the case where the distribution of θ has a distribution function G(θ), the expected utility is

$$\text{EU} = \int [U(C_1(\theta))\rho(\theta) + U(C_2(\theta)+mL)(1-\rho(\theta))]g(\theta)d\theta$$

Hence, the worker is maximizing the following expected utility function by choosing C_1,C_2 and layoff probability (1-ρ): these will be functions of the random price θ that the firm faces, so that the worker wants to maximize their utility given in the last equation subject to the terms that the firm is willing to offer.

What the firm is willing to offer depends on the (expected) profitability of different types of arrangements. Ignoring capital and fixing the number of workers at "n" (ρ fraction of these workers will be employed by the firm, while (1-ρ) of these workers will be on layoff), the value of the firms' output is θf(ρ *n), where f(.) is the production and we have ρ *n employed workers. θ is again the (random) product price. The costs to the firm are

wage costs (for those working)= ρ *n*C_1

compensation costs (for those laid off)= (1-ρ)*n*C_2

so that profits are θf(ρ *n) – ρ *n*C_1 - (1-ρ)*n*C_2. So the expected utility of the firm depends upon its profitability as follows:

EV=

$$\int v(profit(\theta))g(\theta)\,d\theta = \int [v(\theta\ f(\rho*n) - \rho(\theta)*n*C_1(\theta) - (1-\rho(\theta))*n*C_2(\theta)]g(\theta)d\theta$$

where v(.) is the utility the firm receives from its profits, and the other terms are defined as above.

Maximizing the utility of the worker subject to the (expected utility of the) profitability of the firm, we get the first order conditions as follows (suppressing the fact that C_1,C_2, and ρ all are functions of θ under the contract):

1) $U'(C_1) = \lambda n v'(\pi)$
2) $U'(C_2 + m) = \lambda n v'(\pi)$
3) $\rho(1-\rho)[U(C_1) - U(C_2+m) + \lambda\ n\, v'(\pi)(\theta\ f'(\rho n) - C_1 + C_2)] = 0$

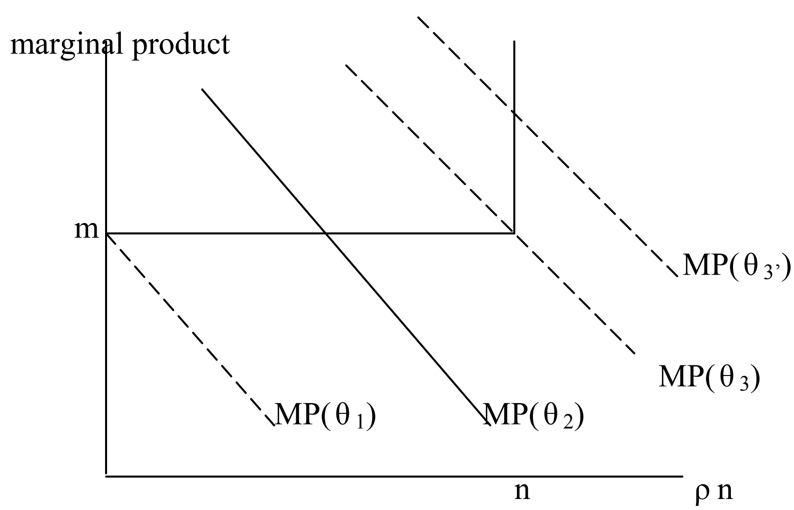
where λ is the Lagrange multiplier and π are profits. Conditions (1) and (2) determine the optimal risk sharing arrangements between agents: marginal utilities are equalized between those laid off and those still working, and marginal utilities of work/nonwork are proportional to all potential outcomes, θ_i. That is, $U'(C_1) = U'(C_2 + m)$ so that $C_1 = C_2 + m$, for any given

θ_i, but C1 and C2 depend on the value of θ_i. For all θ_i, $C_1(\theta) = C_2(\theta) + m$. Hence there is no ex post difference in utility between those laid off and those who aren't for any given θ_i—but those whose firms draw a larger value of θ_i have more desirable outcomes for their employed and laid off workers than firms that drew a lower value of θ_i. Hence, layoffs are voluntary even though those attached to the high-valued θi-firms will be envied by those in the low-valued θ_i-firms. Because of the first two conditions, the third condition becomes:

$$4)\ \rho(1-\rho)[(\theta\, f'(\rho n) - C_1 + C_2)] = 0 \quad \text{or} \quad \rho(1-\rho)[(\theta\, f'(\rho n) - m)] = 0$$

There are three conditions under which this last expression equals zero.
1.If θ is especially low, you will want to shut down the firm so that $\rho = 0$ and $\rho(1-\rho) = 0$, and again the first order condition is satisfied.
2. If $0 < \rho < 1$, then $\theta\, f'(\rho n) = m$, so that marginal product of an additional unit of labor just equals its marginal opportunity cost (giving up the consumption value of leisure, m, when you work full time).
3. If θ is high, the marginal product of labor will be higher than m and you will want to employ more workers, but you have only contracted with n workers. IN this case, $\rho = 1$, so that $\rho(1-\rho) = 0$, and the first order condition is satisfied.

These three cases are illustrated graphically in the following diagram:



Some implications of the (implicit) contracting approach to viewing labor market relations:
a. Contracts are more like marriages, than they are like the "spot" markets where wages decentralize decisions (were firms and workers make unilateral decisions, given wages). Instead, contracts are bilateral agreements made possible by specific HK and (in this case) the desire of workers to insure themselves against random fluctuations in labor demand.
b. Contract resolves the ex-ante uncertainty by finding risk-sharing and income transfer solutions acceptable to both parties.

c. So, contracts involve implicit payments of insurance premiums by workers in good states of the world, and implicit indemnity payments in bad states of the world.