A Short Note on Efficiency Wages

1. Making things a bit more rigorous

In this class we extensively discussed the very fact that what makes labor so hard to regulated by a contract is that effort is not observable by an employer nor is it verifiable. Even if this would not hold, any contract in which wage is a function of some necessarily imperfect measure of effort such as output would allocate too much risk on the worker. However effort is the main actor of the underlying production function not "hours of work traded". A possible way out could be represented by having every agent be a residual claimant or having optimal team contracts. However, being a residual claimant and thus having a wage directly and totally related to output would represent an unbearable risk for agents (which are normally taken to be risk adverse). On the other hand, individual production appears highly unrealistic as economies of scale are pervasive and do play a key role in modern industrial economies inasmuch as they make individual production unrealistic.

The employment relation can be schematically described in the following abstract terms. A principal hires an agent in order to have him perform some value producing actions. The agents submits himself to the principal authority and trades some of his time in return for a wage. This kind of relation is thus potentially beneficial to both the principal (that will be able to appropriate a part of the surplus created by the agent's work) and the agent (that will receive a wage). The point is that once the contract regulating the time/wage has been signed the objects of the two parties do diverge as the principal aims at having the largest possible surplus generated while suffering the disutility of paying a wage to the agent while the agent aims at minimizing the effort he puts in his work given his wage. The key points are that one the one hand the principal cannot observe nor measure the effort exerted by the agent and on the other hand that the relation between effort and output is not perfectly deterministic (i.e. there is a stochastic component in the production function). As said, the real crux of the matter is that the underlying production function has effort as its argument and effort is a non observable quantity. We thus have three main elements so far: surplus, effort and uncertainty: they can be expressed as follows. Let y be the surplus created by the agent's effort e and let us have a stochastic component ϵ(with mean zero) in the functional relation binding y and e. The underlying production function will thus be:

$$y = y(e) + \epsilon$$

Let:

$$e \in [0, 1]$$

It should be noted that while y (output) can be observed, verified and measured effort is not and more to the point, due to the stochastic term in the production function, neither can it be inferred from output itself. Expected utilities will thus be given by:

$$U_P = E[y(e, \epsilon) - w]$$

$$U_A = E[u(w) - c(e)]$$

where $u(w)$ is an increasing and concave function and $c(e)$, representing the cost of effort for the agent, is an increasing and convex function. The kind of situations at stake in our argument are those in which, in addition to the non observability of effort that however remains in our set of assumptions, the contribution of agents to surplus creation (i.e. y) is only imperfectly observable by the principal and hardly verifiable or exceedingly costly verifiable by a third party authority. As said, effort regulation, labor-discipline and efficiency wages models aim at describing this kinds of agency relations (under whose label the employment relation falls). The core of these models is represented by the idea that at least over some range, firms can increase profits by raising wages they pay to some levels above market clearing levels.

The basic tenet is the observation that any agent can either work or shirk: he can either choose an effort level closer to 1 or to 0. As our assumption is that it is costly or right away impossible to control an agent?s behavior, the probability that an agent be found shirking is either zero or negligible. It is noteworthy that here we both have non verifiability as far as surplus creation level is concerned and non observability as far as agents? behavior (choice of effort level) is at stake. Efficiency wages models aims at extracting the maximum amount of labor from hours exchanged by paying an above equilibrium level, fixed and surplus unrelated wage together with a credible threat (usually but not necessarily that of being fired). That is: in a command economy like a firm time is bought and sold, rights are traded for a wage and effort is extracted by incentives, threats and authority.

The main problem for the firm is the determination of an efficiency wage. Should every firm pay the equilibrium wage w^* labor supply and labor demand would clear, there would be no unemployment and every agent would either shirk or chose a zero level of effort until his shirking behavior would be detected and he would be fired. This would be a winning strategy with a low probability of being discovered however, even if fired, the agent would promptly find a new job at the same wage rate (assuming no search nor mobility costs). That is, if $w = w^*$ agent has no incentive to choose a positive effort level as he does not incur any loss in not doing so.

Another strategy for the principal is to set a wage w? at a higher level than the market clearing one. In this circumstance, if fired an agent would lose a wage rent (i.e. the difference between w' and $w*$). The idea is that a sufficiently high wage rent is a strong incentive for the agent to supply high levels of effort (in other terms: a high wage differential is an effective tool for labor extraction). The other firms will however follow in the adoption of this strategy: if each and every firm would be actually paying a wage higher than equilibrium level then a higher equilibrium will be set in which every between-firms wage differential would clear. However another effect would follow: as wages raise, labor demand would reduce and unemployment would raise. In the end: involuntary unemployment would set his role on the stage: his role being that of a credible threat and an instrument of labor discipline.

In such a setting, wages cannot clear the labor market as firms do use wages as an incentive mechanism and would not be willing to hire workers that would accept to work for a lower wage as this would not constitute a credible commitment to effort supply.

If we restrict ourselves to the single period case, we can have agents being characterized by a utility function such as:

$$u = w - e$$

in which net utility is given by wage minus effort. We then have a probability p that the agent is discovered shirking: let this be

$$p \in [0, 1].$$

The firm will pay a wage w^* or fire the employee if found idle. When fired, the employee would obtain a utility w' (which would correspond to an unemployment subsidy or to citizen?s income). If the worker supplies a positive level of effort he gets utility:

$$U_H = w^* - e$$

whereas if he shirks he gets:

$$U_S = (1-p)w^* + pw'$$

as he does not bear any cost but however faces the risk of being fired thus obtaining w'. The efficiency wage is determined on the ground of the inequality

$$U_H \geq U_S$$

that is:

$$w^* - e \geq (1-p)w^* + p\underline{w}$$

from which one easily obtains the efficiency wage:

$$w* = w' + \frac{e}{p}$$

The first key point is that the efficiency wage must be higher than the opportunity cost $w' + e$, that is: the worker must obtain a wage rent, the second point is that the higher the wage the smaller is p. Given all this, firms can either invest in monitoring so to increase p and reduce w^* or increase w^* and reduce the costs of monitoring. In a sense, any decision about this trade off amounts to deciding whether it is the carrot (a higher wage) or the stick (increased monitoring level) that work as an incentive mechanism. As under each of the two perspectives a firm would incur in some costs, it is worthy examining firms decision under the "increasing monitor" and "increasing wages" as well. If we let $c(p)$ be the cost of monitoring per period we can then say that p is the probability that shirking would be detected. In order to minimize costs per period, a firm can choose both wages and monitoring levels as to solve:

$$\underset{p\ w^*}{Min}\{c(p) + w^*\} \text{ such that } p(w^* - w')N \geq z \tag{1}$$

where N is the value of the employment relation in the long run (in other words: the number of "rounds" the interaction is repeated) and z is the benefit that the worker can derive from shirking. The problem for the firm is thus minimizing at each period the costs relative to higher wages and monitoring and the constraint is the existence of the incentive scheme in the second part of the expression. No firm, of course, will ever pay a wage higher than the minimum wage that satisfies the constraint, in the optima solution one thus have:

$$w^* = w' + \frac{z}{Np}$$

and substituting in Eq.1, one obtains:

$$\underset{p\ w^*}{Min}\{c(p) + w' + \frac{z}{Np}.$$

If we let $M(p)$ the cost of monitoring at each period when the probability of detecting shirking behavior is p, then at the optimal value, one must have:

$$M'(p*) - \frac{g}{p^{*^2}N} = 0 \tag{2}$$

$$M''(p*) + 2\frac{g}{p^{*^3}N} \geq 0. \tag{3}$$

According to Eq.2, at the optimal level one must have:

$$M'(p*) = \frac{g}{p^{*^2}N} \tag{4}$$

where the right side measures the change in the efficiency wage level as p increases: this quantity is decreasing in g. The left side measures the increase in the costs of monitoring as

p increases. Note that this quantity can either be increasing or decreasing but, according to expression (5) if it is decreasing it is so at a smaller rate than the right hand side.

Given that wages and control are among the most effective ways to extract labor from labor power, one should ask which of the two is more widely adopted by firms, which of the two is more efficient and under which conditions. Samuel Bowles[1], has stressed that firms tend to prefer costly investments in control (thus increasing p) rather than paying higher wages (given a wage level sufficient to labor extraction). Higher wages are not a waste of resources for society as they are a transfer from employers to employee. Quite on the contrary, control related costs do actually consume real resources and, indeed, the capitalistic firm as an institution, has been described by many as an institution whose main raison d?être is its increased possibility of control rather that superior technological efficiency. If higher wages would be paid and, at the very same time, fewer resources were spent on control those resources could be freed and used for production. The latter proposition is proved by Bowles in the following terms: "The Nash equilibrium resulting from profit maximization by the principal and utility maximization by the agent is both Pareto inefficient and technically inefficient. [...] And taking the Nash equilibrium as the status quo, it would also be possible to revise the employer's labor discipline strategy?reducing monitoring and raising wages, for example?such that the same output could be produced with less of one input (monitoring) and not more of any input."

In a sense, monitoring costs do not fit well with profit maximization and they would not be necessary at all if only levels of effort could be perfectly enforced by a contract. Baker and Hubbard offer a most interesting example. In the eighties, some US trucking companies equipped their trucks with computers with the aim of monitoring drivers' behaviors and actions. These computers allowed companies to have detailed information on a quite remarkable set of divers' and trucks' operations: mostly those in which a conflict of interests was more relevant for the company itself such as speed, idle time and the like. So, for instance, drivers generally prefers to drive faster so they can take longer breaks but the cost of operating trucks is increasing and convex in the speed of the truck. On the other hand, drivers that were also truck owners and thus residual claimants on net revenues internalized all the costs did realized significant savings (and, indeed, they successfully competed with companies' trucks that faced a significant divergence between drivers' and companies' objectives). The key point is that these computers did not provide any improvement – say a more effective coordination between drivers and dispatchers – in the service whatsoever: their main and probably sole function was rather to enlarge the domain of contractibility: the space of drivers' behaviors ? first and foremost those behaviors that more heavily conflicted with companies' interests ? that could be enforced by a contract thanks to making them observable and verifiable. We thus observe a use of technology that by enhancing the possibility of enforceable contracts enhanced profit rates (without producing any other benefit nor improving the service's quality). Key question is: why not adopting a wage increase coupled by a decrease in the costs of monitoring? After all, the same level of effort might have been obtained (according to theory).

2. Keeping things simple

Suppose that an agent can either choose $e = 0$ or $e > 0$ where the latter condition is what we will call *shirking*. The agent's utility function is given by:

$$u = w - e$$

Let us assume less than perfect observability and let us ussume that $p < 1$ where p is the probability that the agent can be caught shirking.

[1] Bowles, S., "The Production Process in a Competitive Economy", *American Economic Review*, 75 (1985), pp.16-36.

The idea is that a firm either pays a positive (higher than equilibrium) wage $w*$ or it fires the agents if this is caught shirking. Once fired, the agent will get a utility $\underline{w}$.

If the agent decides and choose a positive level of effort, his utility will be equal to:

$$U_H = w^* - e$$

whereas if he shirks he gets:

$$U_S = (1-p)w^* + p\underline{w}$$

as he does not bear any cost relative to e but with probability p he is caught shirking thus getting $\underline{w}$.

The efficiency wage is determined as to discourage shirking on the ground of thhe following inequality:

$$U_H \geq U_S$$

that is:

$$w* - e \geq (1-p)w^* + p\underline{w}$$

from which we obtain that

$$w* = \underline{w} + \frac{e}{p}$$

The main result is that the efficiency wage must be larger than the opportunity cost $\underline{w} + e$ (which, in turn, is equal to reservation wage plus effort's disutility). This is what the existence of a quasi wage rent stems from. It is noteworthy that the efficiency wage is higher as much as p is smaller (that is: the harder it is to detect shirking).

So: you choose the stick (invest in control and monitoring and thus make p larger) or the carrot (pay higher wages).