

Shirking, Standards and the Probability of Detection

John G. Sessions* and John D. Skåtun**

**Department of Economics and IZA*

University of Bath

Bath

BA5 7AY

England

Email: j.g.sessions@bath.ac.uk

*** Department of Economics*

University of Aberdeen Business School

Edward Wright Building

Dunbar Street

Old Aberdeen

AB24 3QY

Scotland

Email: j.skatun@abdn.ac.uk

Abstract: By relaxing the common efficiency wage assumption of exogenous shirking detection probabilities, we demonstrate how standards and efficiency wages are related. In a more general setting where the probability of detection depends upon the equilibrium effort level of non-shirkers, we show that the uniformly positive (negative) supply-side relationship between wages (unemployment insurance) and effort is no longer guaranteed. Profit maximization on the part of the firm, however, ensures that effort will depend positively (negatively) on wages (unemployment insurance) in equilibrium.

Word Count: 7051.

Key Words: Monitoring; standards; efficiency wages.

JEL Classification: J33, J41, J54.

Acknowledgements: We are grateful to two anonymous referees and the editors of this journal for helpful comments. The normal disclaimer applies.

26th July 2017

1. Introduction

The fundamental premise of this paper is that efficiency wages and retention standards are related. Whilst the majority of efficiency wage literature has concentrated on imperfect monitoring of effort, we leave monitoring technologies aside and focus instead on the underlying output required by the firm; that is, the minimum *standard* a worker must deliver to continue his employment relationship with the firm. If wages determine effort and effort determines output, and if a firing standard is the critical level of output below which a worker will be fired, then it follows that wages and standards will be related. So, in a world where firms are able to observe worker output but not underlying worker effort, and where output in turn is a function of both luck (i.e. noise) and effort, there must exist a critical level of output below which the worker cannot possibly be exerting the required effort. We take this critical level to be the dividing line between a worker's retention and dismissal. If the worker's output falls short of this standard, then he is fired; if it equals or exceeds the standard, then he is retained and paid the going wage at the firm. It then follows that the higher the wage, the higher the equilibrium level of effort and the higher the standard required of workers.

The central tenet of efficiency wage theory is that wages and effort are positively correlated. In what follows we present a stochastic shirking model in which the robustness of this relationship is tested. If standards, effort and detection probabilities are interdependent, then a sufficiently low standard will ensure that all workers are retained, albeit at low wages and effort. The higher the standard, the higher the required effort and the higher the risk of being detected shirking. By relaxing the literature's common assumption of exogenous shirking detection probabilities and considering instead the more general and endogenous case in which shirking detection depends upon equilibrium effort, we show that the positive

supply-side relationship between efficiency wages and effort is no longer guaranteed.¹ Such a failure may arise when both the cost of effort and the probability of detection are positively correlated with effort but where the former (latter) is positively (negatively) correlated to the wage. This result echoes findings in the monitoring literature and is potentially troubling for efficiency wage theory. However, we demonstrate that in our case this is purely a supply side issue. For when we consider the demand-side we find the efficiency wage being set in a region where the elasticity of the detection probability with respect to effort is less than unity, implying that in equilibrium effort does indeed depend positively on the wage.

In what follows we focus on a fixed wage contract with a set output standard. Workers are dismissed if they breach the contract by failing to perform to the standard set. Whilst this type of contract is common in the labour market, there are alternatives.² These can be broadly divided into contracts where rewards and punishments depend upon either: (i) *absolute performance* - for example, a piece rate in its purest form or other more complicated performance related pay schemes;³ or (ii) *relative performance* – for example, the tournament models espoused originally by Lazear and Rosen (1981)].

We do not focus on the comparative merits of standards versus alternative performance contracts as investigated by, amongst others, Gibbons and Waldmann (1999) and Ghosh and Waldmann (2010). Instead we offer a novel approach that combines standards with efficiency wages. Most efficiency wage models shy away from standards, though there are exceptions. Akerlof (1982), for example, seeks to rationalise why workers would perform beyond a pre-prescribed standard by appealing to gift exchange. We provide a different

¹ Whilst some authors have focussed on the endogenous detection rates that arise through monitoring [see, for example, Calvo and Wellisz (1978) for an early example and Ewing and Payne (1999) for an empirical implementation] we consider instead the endogeneity issues that arise from the choice of effort.

² For a survey on contracts and how they operate in practice, see Malcomson (1997, 1999). For a theoretical exposition of incentives in the labour market and how such incentives affect payment contracts, see Prendercast (1999).

³ The study of piece rates is voluminous and has a long tradition. For early examples see Gibbons (1987) and Lazear (1986).

perspective; by appealing to efficiency wages in the presence of idiosyncratic productivity shocks. In our model we set out the conditions under which non-shirkers will perform at or beyond a set standard, and show how this will at times distinguish them from shirkers.

Conventional efficiency wage theory has traditionally modelled worker effort as the outcome of binary choice decision; workers either shirk by supplying zero effort or they work by exerting the required level of effort. We take a broader view and model effort as a continuum that can be exerted whether working or shirking. We interpret shirking as a neglect of duty by underperforming relative to a required effort level. It encompasses both the conventional zero effort view as well as the more general case of under-exertion in relation to the firm's effort norm.⁴ Our model is therefore closely related to those of Walsh (1999), Goerke (2001), Allgulin and Ellingsen (2002) and Strobl and Walsh (2007), all of whom also assume continuous effort.⁵ By so doing, these authors demonstrate that the trade off between monitoring and wages found in the binary effort dual labour market models of Bulow and Summers (1986) does not automatically transfer to the case where effort is continuous. We differ from these authors, however, by concentrating on standards rather than monitoring. Our approach opens up a new series of results linking standards to effort and the probability of detection. Thus, we demonstrate that shirking declines when standards are raised. Intuitively, higher standards increase the probability of detection for a given level of effort. As a result, shirkers, who optimise their trade-off between the cost of effort and the risk of being identified, increase their effort in response to the increase in detection probability.

A question worth asking is under what circumstances will a required minimum standard be the appropriate mechanism to trigger dismissal. Is it not possible that firms could

⁴ Our focus is on standards rather than the psychological norms discussed in relation to unemployment and the labour market by Akerlof (1980) and Clark (2003).

⁵ Hahm and Mayer (2011) show in a model of efficiency wages and search that even when effort is binary it is possible that monitoring (or detection rates) and wages are not necessarily substitutes.

use relative performance to dismiss employees? This would certainly appear to be the usual practice in team sports where athletes who underperform relative to their teammates face termination of their contracts. It follows from this that an alternative modelling strategy, in which relative performance plays a more prominent role, might be to adapt a tournament model in the Lazear and Rosen (1981) tradition to incentivise workers through punishment rather than reward; i.e. where relatively poor performing workers are sacked. From an economic perspective it makes sense for the firm to evoke such a firing trigger strategy if performance is noisy but all employers face the same aggregate unobserved shock. On the other hand, if shocks are idiosyncratic rather than common, as in our model, then tournaments are dominated by contracts as initially demonstrated by Green and Stokey (1982). Our paper therefore does not follow the tournament route but is more akin to the literature on standards or thresholds as part of incentive schemes. Although these have been typically ignored in the efficiency wage literature, they have a long tradition elsewhere. For instance, an early exposition by Mirrlees (1974) investigates how it might be optimal to punish agents who do not attain a given performance threshold. There is also a growing literature relating to bonuses, emanating from Healy (1985), in which CEO's seek to shift earnings (i.e. output performance) to later periods whenever performance exceeds an upper threshold.⁶

It is apparent from the above discussion that whether relative or absolute performance is the chosen measurement criteria for dismissal depends on the nature of shocks and the underlying economic reasoning. However, it is not the whole story since legal frameworks and employment law may also play important roles. There are varying limitations across different judicial regions on when you can fire workers - see, for instance, Blau and Kahn (1999) for a discussion of employment-protection legislation that makes it costly or difficult

⁶ For an exposition on how bonuses relate to two thresholds - a lower one where bonuses kick in and a higher one where bonuses are capped - see Murphy (1999, 2013) and Murphy and Jensen (2011).

for employers to terminate jobs without cause. If such legislation is enforced and absolute standards of performance are used in court or in industrial tribunals, then these legal restrictions may suggest that absolute standards in firing may be more appropriate than relative performance criteria to determine dismissals. Thus, absolute performance measures may be particularly applicable to legal jurisdictions with extensive employment protection, such as the original member countries of the European Union. The legal argument for absolute performance may apply to a lesser extent in the United States where the general rule is that firms have the right to fire at will, although even here unjust firing laws exists.

We proceed in Section 2 to develop a model which maintains an absolute standard of acceptable worker related output, below which workers are fired, to investigate the interdependent effects of standards, effort, wages and the probability of shirker detection. In Section 3 we conclude.

2. The Model

We present an efficiency wage model in the Shapiro and Stiglitz (1984) tradition that is extended to include a stochastic element. We consider the case where the firm observes worker output but where output is a function of both worker effort and an idiosyncratic stochastic shock. Workers are retained and paid the efficiency wage providing their observed output does not fall below a defined standard. Given our focus of endogenising the probability of detecting shirking behaviour, we do not model the unemployment consequences of the efficiency wage on the wider labour market.

To conceptualise the informational context of our model it is helpful to consider the time sequence for each period t illustrated in Figure 1 following:

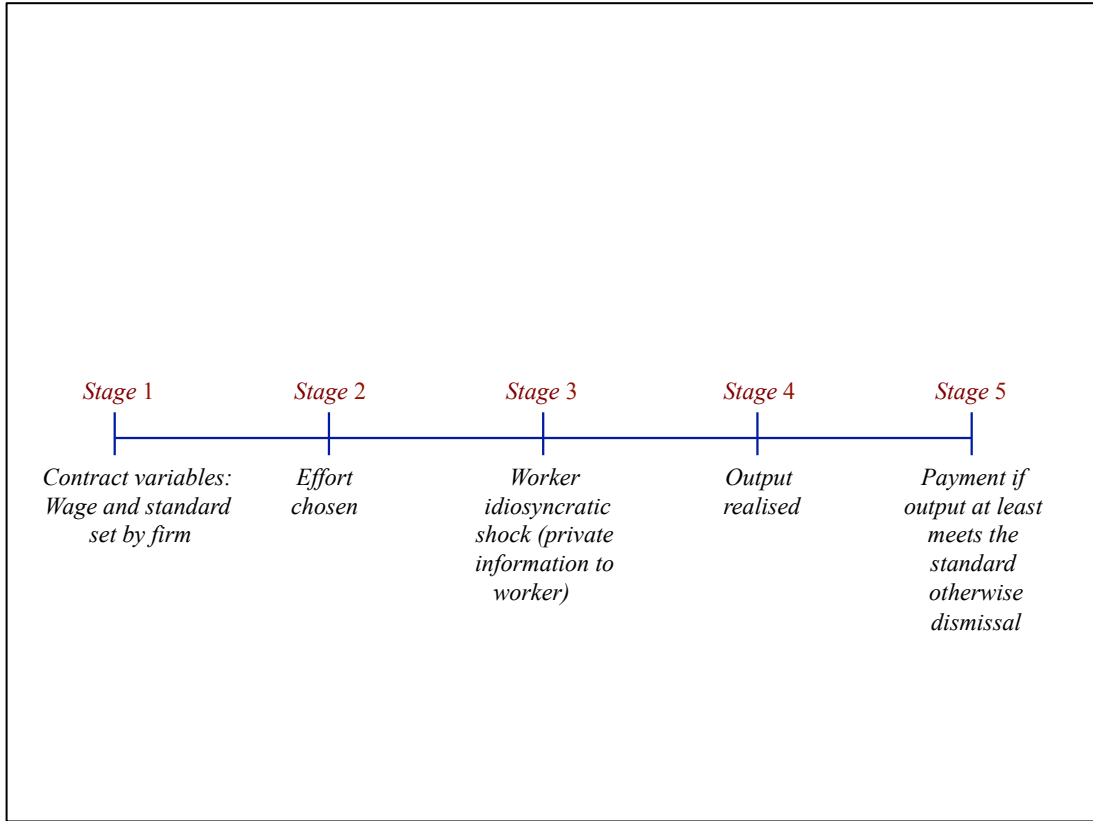


Figure 1: The Informational Context in Each Time Period

In Stage 1, for each period t , the contract variables are set by the firm vis. the wage and standard. In Stage 2 an effort level is chosen by the worker but is not necessarily known by the firm. After effort is chosen the worker experiences an idiosyncratic shock to his productivity in Stage 3, which the firm cannot observe. Output, however, is realised as and is common knowledge in Stage 4. Finally, in Stage 5, a worker is remunerated if output at least equals the standard and is dismissed otherwise.

Formally, we assume that workers are identical, risk neutral and endowed with a separable utility function, $u(w, e) = w - c(e)$, where w and $e \in [0, e^+]$ denote income and worker effort respectively and where $c(\cdot)$ is a continuous and convex cost function with $dc(e)/de \equiv c'(e) > 0$, $d^2c(e)/de^2 \equiv c''(e) > 0$ and $c(0) = c'(0) = 0$.⁷ Each worker is associated with a stochastic output function, $y_i = \theta_i f(e)$, which varies with state i . Workers

⁷ Note that the function $c(e) = Ae^\alpha$, where $\alpha > 1$, which naturally embeds the quadratic cost function, satisfies these conditions.

choose effort prior to the realisation of this output shock and technology is such that $df(e)/de \equiv f'(e) > 0$, $d^2f(e)/de^2 \equiv f''(e) < 0$, $f(0) = 0$ and $f'(0) = \infty$. The shift-parameter, θ_i , represents a random shock to productivity in state i and is uniformly distributed between θ_L and $\theta_H > \theta_L$. For an individual worker θ_i reflects relative misfortune (when it is low) or luck (when it is high).

The firm's objective is to maximise per-worker expected profit, $\pi = \mathbf{E}\{\theta_i f(e)\} - w$, subject to providing the worker with at least his outside option (e.g. unemployment insurance) utility $b < w$.⁸ Writing this participation constraint as:

$$\begin{aligned} u(w^r, e) &= w^r - c(e) = b \\ \Rightarrow \\ w^r &= w^r(b, e) = c(e) + b \end{aligned} \tag{1}$$

yields an inverse function in which the reservation wage, w^r , depends on effort and the outside option. The nature of this relationship is ascertained from totally differentiating expression (1) vis: $dw^r/de \equiv w'_e(b, e) = c'(e) > 0$; $d^2w^r/de^2 \equiv w''_{ee}(b, e) = c''(e) > 0$; $dw^r/db \equiv w'_b(b, e) = 1 > 0$ and $d^2w^r/db^2 \equiv w''_{bb}(b, e) = 0$. Thus, the firm's profit maximising level of effort, e^* , is defined implicitly from:

$$\frac{\partial \pi(b, e)}{\partial e} = \pi_e(b, e^*) = \theta_i f'(e^*) - w'_e(b, e^*) = 0 \tag{2}$$

The problem facing the firm is that whilst it is able to observe worker output, it is unable to observe either worker effort, e , or 'luck', θ_i . Nonetheless, it seems reasonable to assume that in some instances effort can be partially deduced. To reflect this, consider the case where the firm sets a 'standard'; that is, a minimum level of output, \tilde{y} , that the worker must attain in

⁸ Note that we have normalised the product price to equal unity for the sake of simplicity.

order to be retained in the workplace, determined by the lowest possible output produced by a non-shirking worker.⁹ We define a critical realisation to a shirker of the random shock, $\tilde{\theta}$, below which shirking (i.e. supplying less than required effort) will always be detected. Formally, we assume:

$$\tilde{y} = \theta_L f(e^*) = \tilde{\theta} f(\bar{e}) \quad (3)$$

where e^* denotes the firm's choice level of effort and $\bar{e} < e^*$ denotes the 'shirking' level of effort. It thus follows that the worst case scenario when the worker supplies the firm's desired level of effort in the least favourable state of nature defines implicitly a critical state of nature at which anything less than required effort will be detected. The critical state therefore satisfies:

$$\tilde{\theta} \equiv \tilde{\theta}(\bar{e}, e^*) = \frac{\theta_L f(e^*)}{f(\bar{e})} \quad (4)$$

It is apparent that the critical state is increasing in the firm's desired level of effort and decreasing in shirking effort:

$$\frac{\partial \tilde{\theta}}{\partial e^*} \equiv \tilde{\theta}_{e^*}(\bar{e}, e^*) = \frac{\theta_L f'(e^*)}{f(\bar{e})} > 0 \quad (5)$$

$$\frac{\partial \tilde{\theta}}{\partial \bar{e}} \equiv \tilde{\theta}_{\bar{e}}(\bar{e}, e^*) = -\frac{\theta_L f(e^*) f'(\bar{e})}{[f(\bar{e})]^2} < 0 \quad (6)$$

Intuitively, a higher desired level of effort, e^* , on the part of the firm raises the acceptable bar of output performance (i.e. the standard \tilde{y}) resulting in more states in which shirking is

⁹ Thus, no non-shirking worker will be fired though some shirking workers could survive. Proposition 1 and Corollary 1 to follow, determine the conditions for when setting such a standard is optimal for the firm.

identifiable. Thus, the temptation to shirk declines with the equilibrium level of non-shirking effort as potential shirkers can expect to be detected more frequently and must hence hope for a higher realisation of luck to avoid being dismissed.

These assumptions are represented in Figure 2 and Figure 3 following. Recalling that both a shirker and a non-shirker choose their effort level prior to the realisation of the state of the world, the figures reflect possible output levels. The two upward sloping lines in Figure 2 depict the outputs generated by a shirking worker, $\bar{y}_i = \theta_i f(\bar{e})$, and a non-shirking worker, $y_i^* = \theta_i f(e^*)$. Since the firm is only able to observe output, but not its constituent elements (i.e. effort and luck), it is unable to distinguish between a shirker whose productivity realisation is $\tilde{\theta}(\bar{e}, e^*)$ and a non-shirker whose productivity realisation is at the lower bound θ_L . More generally, the firm is unable to detect shirking at any productivity realisation $\theta_i \geq \tilde{\theta}(\bar{e}, e^*)$. Shirking is, however, detectable at any productivity realisation $\theta_i < \tilde{\theta}(\bar{e}, e^*)$ since the revenue from shirking here falls short of the lowest output possible for a non-shirker.

In terms of Figure 3, an increase in the equilibrium non-shirking level of effort from e_0^* to $e_1^* > e_0^*$, which is equivalent to an increase in the standard from \tilde{y}_0 to $\tilde{y}_1 > \tilde{y}_0$, increases the critical shift parameter from $\tilde{\theta}(\bar{e}, e_0^*)$ to $\tilde{\theta}(\bar{e}, e_1^*) > \tilde{\theta}(\bar{e}, e_0^*)$. Thus, as the lowest possible output for non-shirkers increases, potential shirkers need to be even luckier to avoid detection.

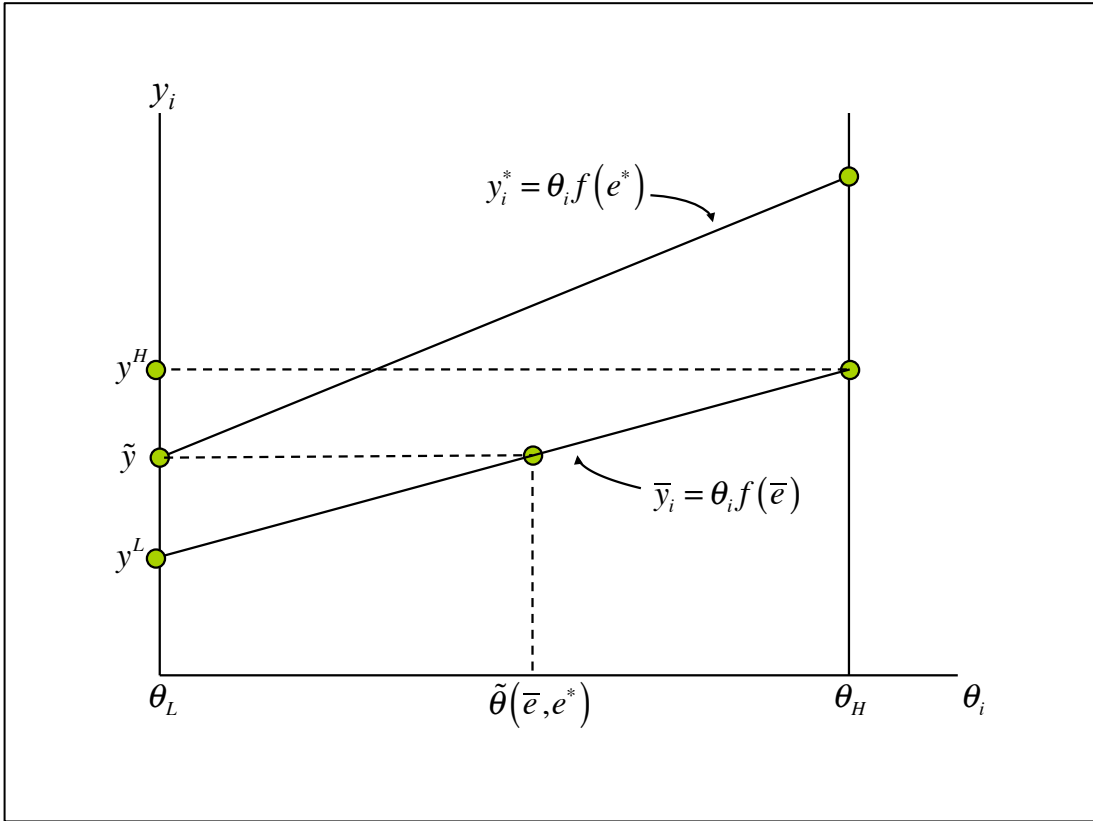


Figure 2: Critical 'Luck'

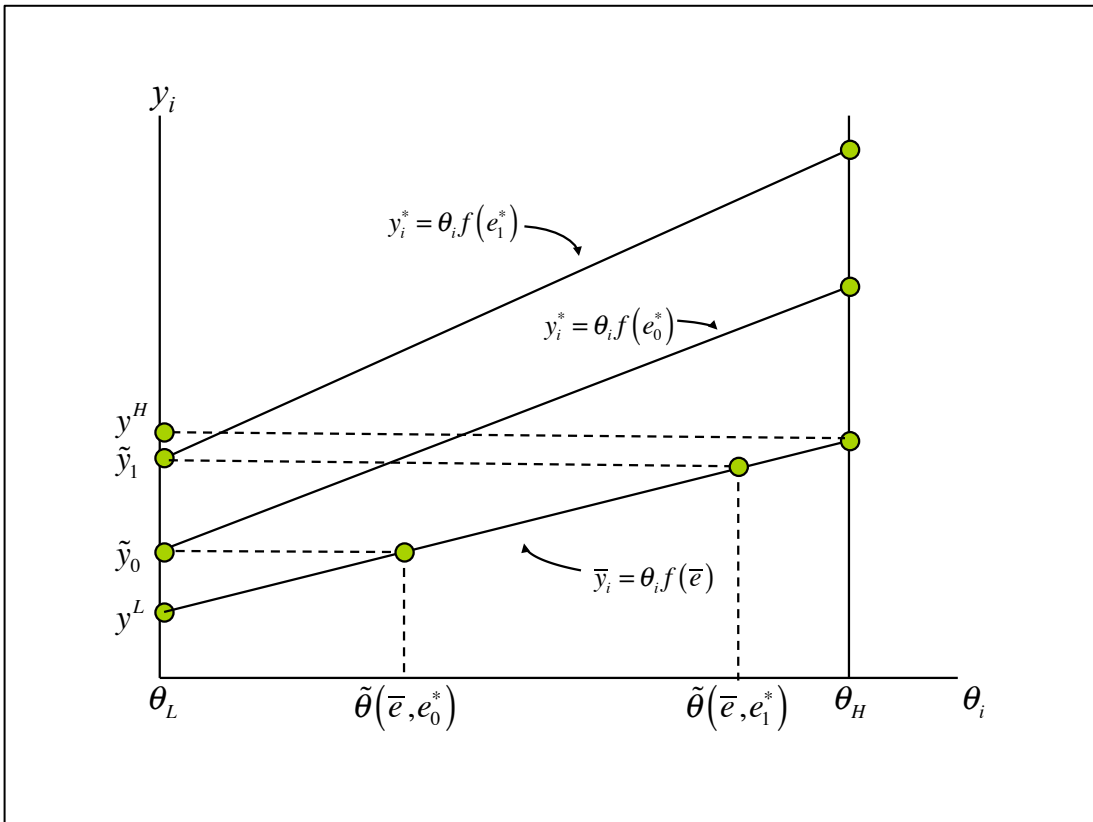


Figure 3: Critical 'Luck' and Effort

Let us depart for the time being from the case where the standard is set and restricted by (3), and instead consider the case where the firm can set any standard y , where y does not necessarily equal \tilde{y} . In this context, consider the worker's decision problem over effort vis. supplying the effort required to attain the standard or supplying a lower (i.e. shirking) level of effort. We now allow for both type-1 and type-2 errors such that shirking (non-shirking) workers may be inadvertently retained (fired). The supply of effort from non-shirking workers will be determined by an incentive compatible 'non-shirking constraint' (NSC). This specifies the lowest wage a worker will accept in return for supplying a given level of effort or, equivalently, the maximum effort the worker will supply for a given wage. Intuitively, workers will provide the firm's required level of effort, e^* , if the expected utility from so doing is at least as great as that from shirking. The NSC is thus:

$$(1 - \hat{p})w + \hat{p}b - c(e^*) \geq pb + (1 - p)w - c(\bar{e}) \quad (7)$$

where p and \hat{p} denote respectively the probabilities that a shirker and non-shirker are fired.

The dismissal probability for a non-shirker given a standard y is:

$$\hat{p} = \begin{cases} 0 & \text{if } y \leq \tilde{y} \\ \frac{\theta_n^c - \theta_L}{\theta_H - \theta_L} & \text{otherwise} \end{cases} \quad (8)$$

where $\theta_n^c = y/f(e^*)$ denotes the critical state at which non-shirkers are fired given the standard y . The dismissal probability for a shirker is given by:

$$p = \begin{cases} 0 & \text{if } y \leq y^L \\ \frac{\theta_s^c - \theta_L}{\theta_H - \theta_L} & \text{otherwise} \\ 1 & \text{if } y \geq y^H \end{cases} \quad (9)$$

where $y^L = \theta_L f(\bar{e})$, $y^H = \theta_H f(\bar{e})$ and $\theta_s^c = y/f(\bar{e})$ denotes the critical state at which non-shirkers are fired given a standard y .

Satisfaction of the NSC implies an incentive compatible (i.e. efficiency) wage schedule:

$$w^* = b + \left[\frac{c(e^*) - c(\bar{e})}{p - \hat{p}} \right] \quad (10)$$

The efficiency wage, w^* , is the lowest wage compatible with the provision of a given level of non-shirking effort e^* . It is increasing in the worker's outside unemployment opportunity, b , and effort cost, $c(\cdot)$, since the firm will have to pay more to induce effort when alternative employment prospects are good and when the supply of effort is more onerous. The wage is also increasing in the probability, \hat{p} , of high effort workers being inadvertently sacked. In contrast, the wage is decreasing in the probability, p , of the correct detection of shirkers with workers becoming more wary of shirking as the risk of detection increases. Since both of these probabilities are functions of the standard, a potential trade-off emerges.

The trade-off can be seen graphically in Figure 2. Should the firm choose to set a standard $y = \tilde{y}$ then it would retain all non-shirkers whilst unlucky shirkers would be fired. If the standard is set above \tilde{y} such that $y > \tilde{y}$ then the probability of shirkers being fired increases. The shirkers would in essence need to ride their luck more often. Thus the firm would not retain the low effort workers as often. However, the firm would now start making firing mistakes in the sense that it would on occasion fire the wrong type of worker; the high effort non-shirkers. Whilst there might be an advantage in raising the standard above \tilde{y} to punish the shirkers more often, there is an associated cost of unfairly punishing hard working employees. There is no advantage of increasing the standard above y^H as all shirkers will

then have already been fired. There is on the other hand no point in having a standard below \tilde{y} as the firm would simply be punishing the shirkers less often whilst retaining all non-shirkers. Thus the range of operative standards must be $y \in [\tilde{y}, y_H)$. Clearly, no standards will be set outside this range.

It is apparent from Figure 3 that the range of possible standards $[\tilde{y}, y_H)$ is a function of the non-shirkers effort level. Consider the case where we for simplicity assume a fixed level of shirking effort such that the output idiosyncratic shock level (the upward sloping curve) of shirkers remains fixed. Now let the non-shirkers experience an exogenous change in their effort level such that their upward sloping output-luck locus shifts leftward and up. At any given luck realisation the non-shirkers output increases. This implies that shirkers will on average have to be luckier to be retained by the firm since the critical output \tilde{y} increases; $\tilde{y}: \tilde{y}_0 \rightarrow \tilde{y}_1$ with higher effort. Thus the range of possible standards shrinks.

Assuming that the firm sets a standard $\tilde{y} = \theta_L f(e^*)$, such that only shirking workers are fired, then the probability of a shirker being detected and fired is given by:

$$p[\tilde{\theta}(\bar{e}, e^*)] \equiv \frac{[\tilde{y}/f(\bar{e})] - \theta_L}{\theta_H - \theta_L} = \frac{\tilde{\theta}(\bar{e}, e^*) - \theta_L}{\theta_H - \theta_L} = \frac{\theta_L}{\theta_H - \theta_L} \left[\frac{f(e^*)}{f(\bar{e})} - 1 \right] \quad (11)$$

In this case, the worker's decision problem regarding effort is given by:

$$\max_{\bar{e}} \mathbf{E}\{u(w, \bar{e})\} = p[\tilde{\theta}(\bar{e}, e^*)]b + \{1 - p[\tilde{\theta}(\bar{e}, e^*)]\}w - c(\bar{e}) \quad (12)$$

where:

$$p[\tilde{\theta}(\bar{e}, e^*)] = \begin{cases} 0 & \text{if } \bar{e} \geq e^* \\ \frac{\theta_L}{\theta_H - \theta_L} \left[\frac{f(e^*)}{f(\bar{e})} - 1 \right] & \text{otherwise} \end{cases} \quad (13)$$

Clearly, there will either be an interior (i.e. shirking) or corner (i.e. non-shirking) solution to this maximisation problem. If the former, then the worker's optimal choice of effort, \bar{e} , is derived implicitly from the first order condition:

$$\begin{aligned} \frac{\partial \mathbf{E}\{u(w, \bar{e})\}}{\partial \bar{e}} &= \frac{f'(\bar{e})}{f(\bar{e})^2} \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] (w - b) - c'(\bar{e}) = 0 \\ \Rightarrow & \\ \frac{f'(\bar{e})}{f(\bar{e})^2} \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] (w - b) &= c'(\bar{e}) \end{aligned} \tag{14}$$

Intuitively, a potential shirker will provide effort up to the point at which the marginal benefit from so doing, namely the reduction in the probability of losing the rent of wages over unemployment insurance, equals the marginal cost of increasing effort.

Assuming $\tilde{y} = \theta_L f(e^*)$, then the worker's NSC reduces to:

$$w - c(e^*) \geq p[\tilde{\theta}(\bar{e}, e^*)]b + \{1 - p[\tilde{\theta}(\bar{e}, e^*)]\}w - c(\bar{e}) \tag{15}$$

Satisfaction of the reduced NSC implies an incentive compatible (i.e. efficiency) wage schedule:

$$w^* = b + p[\tilde{\theta}(\bar{e}, e^*)]^{-1} [c(e^*) - c(\bar{e})] \tag{16}$$

The efficiency wage, w^* , is the lowest wage compatible with the provision of a given level of non-shirking effort e^* (i.e. the standard \tilde{y}). It follows a similar intuition to expression (10). For instance, higher detection probabilities of shirkers still shade the necessary effort-inducing wage that the firm is obliged to offer.

Increasing the standard beyond the point at which only the shirkers are sacked to regions where non-shirkers also face unemployment risk will have an a priori ambiguous

effect on wages. However, consider the conditions under which the firm will not find it optimal to diverge from a standard where only shirkers are sacked. This occurs when non-shirkers are always retained. Formally:

Proposition 1: *The firm will find it optimal to set the standard $\tilde{y} = \theta_L f(e^*)$ iff $\partial p/\partial y > 0$, $\forall y$ and $(\partial p/\partial y) - (\partial \hat{p}/\partial y) < 0$, $\forall y > \tilde{y}$.*

Proof: *From expression (10) above it follows that the firm's expected profit can be written:*

$$\mathbf{E}\{\pi\} = \mathbf{E}\{\theta_i f(e^*)\} - w^* = \mathbf{E}\{\theta_i f(e^*)\} - b - \left[\frac{c(e^*) - c(\bar{e})}{p - \hat{p}} \right] \quad (17)$$

The firm will choose effort e^ and the standard y to maximise profit.¹⁰ The latter choice variable yields the following first order total derivative:*

$$\frac{d\pi}{dy} = \frac{[(\partial p/\partial y) - (\partial \hat{p}/\partial y)][c(e^*) - c(\bar{e})]}{(p - \hat{p})^2} + \frac{\partial \pi}{\partial e^*} \cdot \frac{\partial e^*}{\partial y} \quad (18)$$

where $\partial \pi/\partial e^ = 0$ from expression (2). It therefore follows that $\text{sgn}(d\pi/dy) = \text{sgn}[(\partial p/\partial y) - (\partial \hat{p}/\partial y)]$. From expression (8) we have $\partial \hat{p}/\partial y > 0$ if $y \leq \tilde{y}$. In addition, if $(\partial p/\partial y) - (\partial \hat{p}/\partial y) < 0$, $\forall y > \tilde{y}$ then:*

$$\text{sgn}\left(\frac{d\pi}{dy}\right) = \begin{cases} > 0 & \text{if } y < \tilde{y} \\ < 0 & \text{if } y > \tilde{y} \end{cases} \quad (19)$$

With a continuous profit function it thus follows that expected profit is maximised at $y = \tilde{y}$.

QED.

Proposition 1's dependence on $(\partial p/\partial y) - (\partial \hat{p}/\partial y)$ may be given an intuitive explanation.

Consider the efficiency wage schedule set out in expression (10) previously, where it follows that if the sign of $(\partial p/\partial y) - (\partial \hat{p}/\partial y)$ is positive (negative) then an increase in the standard

¹⁰ Note that the firm is therefore also implicitly choosing the probabilities of being laid off as well as the efficiency wage.

will reduce (increase) the efficiency wage needed to ensure incentive compatibility. With the conditions of Proposition 1 it follows that when $y > \tilde{y}$, an increase in the standard will necessitate a higher efficiency wage such that the firm's profit is declining in the standard. Conversely, an increase in the standard when $y < \tilde{y}$ will reduce the efficiency wage required to induce non-shirking such that the firm's profit is increasing in the standard. It therefore follows that an increase in the standard will increase profits when $y < \tilde{y}$ and decrease profits when $y > \tilde{y}$. It thus follows that the firm's profits are maximised when the standard is set at $y = \tilde{y}$.

That is not all. Further insight follows from Proposition 1 by considering the elasticities of effort with respect to the standard for shirkers and non-shirkers:

Corollary 1: *The firm will find it optimal to set the standard $\tilde{y} = \theta_L f(e^*)$ iff ε_n is sufficiently small in comparison to ε_s , where $\varepsilon_n \equiv (\partial e^*/\partial y)(y/e^*)$ and $\varepsilon_s \equiv (\partial \bar{e}/\partial y)(y/\bar{e})$ denote respectively the elasticity of non-shirking effort and shirking effort with respect to the standard.*

Proof:

Note:

$$\frac{\partial p}{\partial y} = \frac{1}{\theta_H - \theta_L} \left\{ \frac{f(\bar{e}) - \bar{e}f'(\bar{e})\varepsilon_s}{[f(\bar{e})]^2} \right\} \quad \forall y \in (y_L, y_H) \quad (20)$$

Thus, $\partial p/\partial y > 0$ iff $\varepsilon_s \leq 1$. Similarly:

$$\frac{\partial \hat{p}}{\partial y} = \begin{cases} 0 & \text{if } y \leq \tilde{y} \\ \frac{1}{\theta_H - \theta_L} \left\{ \frac{f(e^*) - e^*f'(e^*)\varepsilon_n}{[f(e^*)]^2} \right\} & \text{otherwise} \end{cases} \quad (21)$$

Thus, $\partial \hat{p}/\partial y > 0$ if $\varepsilon_s \leq 1$. Corollary 1 now follows directly from Proposition 1.

QED

Proposition 1 puts in essence restrictions on how responsive in terms of effort shirkers are to standards as compared to non-shirkers. Proposition 1 intuitively holds as long shirkers' effort is sufficiently responsive to standards. We will henceforth assume that Proposition 1 and Corollary 1 hold such that the firm sets the standard as determined by expression (4).

Assuming then that the firm sets a standard $\tilde{y} = \theta_L f(e^*)$ such that only shirking workers are fired, then the probability of a shirker being detected and fired is therefore determined by expression (11). In contrast to the conventional efficiency wage story, this probability is determined endogenously by the equilibrium level of effort. Indeed, we now derive:

Proposition 2: *The probability of detecting shirking depends positively on the equilibrium effort level of non-shirkers.*

Proof: *Partial differentiation of expression (11) above yields*

$$\partial p / \partial e^* = \tilde{\theta}_{e^*}(\bar{e}, e^*) / (\theta_H - \theta_L) = \theta_L f'(e^*) / f(\bar{e})(\theta_H - \theta_L) > 0.$$

QED

As Proposition 2 states and Figure 3 illustrates, the probability of detecting (and thus dismissing) a shirker increases with equilibrium effort since this raises the critical shift parameter, leaving the transgressor less states in which to hide. That is, workers who raise their effort level to the gratification of firms do so to the detriment of potential shirkers who are more readily identifiable. Proposition 2 thus stands in sharp contrast to previous literature in which effort and detection probabilities are unrelated.¹¹ As equilibrium effort effectively determines the critical dismissal-retention output, \tilde{y} , we can also draw inferences between standards and the probability of detection. Thus, within an efficiency wage framework we find that increasing standards increases the probability of detection. This echoes the findings of Rasmusen and Zenger (1990) who, using a teamwork model of agency in the Holmstrom

¹¹ Though not directly linked, Proposition 1 suggests a fair amount of introspection with respect to effort in relation to internal effort levels within the firm, not dissimilar to the discussion in Akerlof and Yellen (1990) and Danthine and Kurmann (2009) where the central theme is the relative wage within the firm as opposed to an external reference wage.

(1982) tradition, demonstrate that the probability of detecting shirking increases with the output target set.

The expected utility from shirking, which is detected only if $\theta < \tilde{\theta}$, is given by:

$$u(\bar{e}, e^*) = pb + (1-p)w - c(\bar{e}) \quad (22)$$

It can be shown that shirking workers will never provide zero effort. To be sure:

Proposition 3: *A shirking worker will operate in the region $\bar{e} \in (0, e^*)$.*

Proof: *See Appendix.*

If the probability of detecting shirking is endogenous then it follows that shirkers will not necessarily exert zero effort, as is commonly assumed in the efficiency wage literature. Whilst shirkers by definition exert less effort than that required by the firm, they trade off the cost of effort against the reduction in the detection probability and do best by exerting at least some effort.

Note that if the standard is set sufficiently high, or the wage sufficiently low, then all (identical) workers will shirk - in the sense that they fail to provide the level of effort consistent with always attaining the standard set by the firm. They then will all run the risk of being fired. In this case we note:

Proposition 4: *When all workers shirk an increase in the wage will increase shirking effort such that $d\bar{e}/dw > 0$.*

Proof: *This can be demonstrated by totally differentiating the first-order utility maximising condition (given in the proof of Proposition 3) with respect to wages and effort.*

$$\begin{aligned} \frac{d\bar{e}}{dw} &= \frac{\frac{\partial p}{\partial \bar{e}}}{\frac{\partial^2 p}{\partial \bar{e}^2}(b-w) - c''(\bar{e})} \\ \Rightarrow & \\ \frac{d\bar{e}}{dw} &= \frac{f'(\bar{e})f(\bar{e})}{\left[f''(\bar{e})f(\bar{e}) - 2f'(\bar{e})^2 \right] (b-w) + c''(\bar{e})f(\bar{e})^3 \left[\frac{(\theta_H - \theta_L)}{\theta_L f(e^*)} \right]} > 0 \end{aligned} \quad (23)$$

since:

$$\frac{\partial p}{\partial \bar{e}} = -\frac{f'(\bar{e})}{f(\bar{e})^2} \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] < 0 \quad (24)$$

And:

$$\frac{\partial^2 p}{\partial \bar{e}^2} = -\left[\frac{f''(\bar{e})f(\bar{e}) - 2f'(\bar{e})^2}{f(\bar{e})^3} \right] \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] > 0 \quad (25)$$

QED.

The conventional efficiency wage result that wages and effort are positively correlated is retained. Higher wages increase the fear of dismissal and induce shirkers to raise effort, albeit not necessarily to the required standard.

We are now able to draw inferences as regards how shirkers react to standards within the firm and outside opportunities.

Proposition 5: *Shirkers will exert: (a) more effort the higher the standard, \tilde{y} (as reflected by a higher e^*) set by the firm; and (b) less effort the higher the outside option utility, b :*

Proof: *Part (a) can be demonstrated by totally differentiating the first-order utility maximising condition (given in the proof of Proposition 3) with respect to shirking effort, \bar{e} , and standard (i.e. non-shirking) effort, e^* :*

$$\begin{aligned} \frac{d\bar{e}}{de^*} &= - \left[\frac{\frac{\partial^2 p}{\partial \bar{e} \partial e^*} (b-w)}{\frac{\partial^2 p}{\partial \bar{e}^2} (b-w) - c''(\bar{e})} \right] \\ &\Rightarrow \\ \frac{d\bar{e}}{de^*} &= - \left[\frac{f'(e^*) f'(\bar{e}) f(\bar{e}) \theta_L (b-w)}{\Phi \theta_L f(e^*) (b-w) + c''(\bar{e}) f(\bar{e})^3 (\theta_H - \theta_L)} \right] > 0 \end{aligned} \quad (26)$$

where $\Phi = f''(\bar{e}) f(\bar{e}) - 2f'(\bar{e})^2$. Part (b) can similarly be proven by differentiating this condition with respect to shirking effort, \bar{e} , and outside option utility, b , yielding:

$$\begin{aligned} \frac{d\bar{e}}{db} &= - \left[\frac{\frac{\partial p}{\partial \bar{e}}}{\frac{\partial^2 p}{\partial \bar{e}^2} (b-w) - c''(\bar{e})} \right] \\ &\Rightarrow \\ \frac{d\bar{e}}{db} &= - \left\{ \frac{f'(\bar{e}) f(\bar{e})}{\Phi (b-w) + c''(\bar{e}) f(\bar{e})^3 \left[\frac{(\theta_H - \theta_L)}{\theta_L f(e^*)} \right]} \right\} < 0 \end{aligned} \quad (27)$$

QED.

Proposition 5 reflects the considerations a potential shirker makes with respect to the possibility of being detected and fired, and so forfeiting wages in exchange for unemployment utility. Increasing the standard, \tilde{y} (i.e. raising the required non-shirking effort level e^*), as in part (a), is equivalent to the firm becoming less tolerant as regards low output. Thus, the probability of a shirker being detected is effectively increased as result of the firm's higher standards. To countervail this effect, the shirker responds by increasing effort. The penalty of being detected is simply the difference between the wage if employed and unemployment utility if fired. Any increase in the latter, as in part (b), will have an adverse effect on effort. This is a common result in the traditional shirking literature, where typically

no one shirks in equilibrium. The novel aspect here is that this result translates into a situation where some or all workers shirk.¹²

Returning to expression (16), it follows that since both the cost of effort and the probability of detection of shirkers are positively correlated with effort but oppositely (i.e. cost of effort - positively; probability of detection – negatively) correlated with the wage, that a new complexity has arisen whereby the relationship between the incentive compatible wage and effort is not unambiguously positive. To be sure:

Proposition 6: *A sufficient (but not necessary) condition for the positive supply-side correlation between the (efficiency) wage and (non-shirking) effort (i.e. $\partial e^*/\partial w^* > 0$) is that the elasticity of the probability of detection of shirkers with respect to non-shirking effort, η^* , is less than unity.*

Proof: *From expression (16) it follows that:*

$$\frac{\partial e^*}{\partial w^*} = \frac{p}{c'(e^*) - [c(e^*) - c(\bar{e})] \frac{\partial p}{\partial e^*} \frac{1}{p}} = \frac{pe^*}{c'(e^*)e^* - \eta^* [c(e^*) - c(\bar{e})]} \quad (28)$$

where $\eta^* = (\partial p / \partial e^*)(e^*/p) > 0$ denotes the elasticity of the probability of detection with respect to effort and $p \equiv p[\tilde{\theta}(\bar{e}, e^*)]$. Since $c'(e^*)e^* - c(e^*) > 0$ by the convexity of the cost function, the proposition follows.¹³

QED.

Proposition 6 illustrates a potential fissure in the positive link between efficiency wages and effort. Only by constraining the effect of effort on the probability of detection to be relatively small as compared to the effect of effort on the worker's cost (i.e. disutility of effort), are we

¹² This is in the tradition of Shapiro and Stiglitz (1984) - hereafter SS - where no worker shirks in equilibrium. However, unlike SS – see following - we make no inference about the involuntary unemployment implications of the efficiency wage w^* : 'From the worker's point of view, unemployment is involuntary: those without jobs would be happy to work at w^* or lower, but cannot make a credible promise not to shirk at such wages.' [SS (1984), p. 438]. Whilst we recognise that the higher (non-shirking) wage in our model may induce involuntary unemployment, as the unemployed workers outside the firm would prefer to work at the firm, it is not the focus of our paper.

¹³ Proposition 5 contains a sufficiency but not a necessary requirement since effort may rise with the wage even if $\eta^* > 1$ when $c(\cdot)$ is sufficiently convex or when the difference between shirking and non-shirking effort is sufficiently small.

able to retain the intuitively attractive positive correlation between the supply of effort and wages. This condition resembles those in Walsh (1999) and Strobl and Walsh (2007), both of whom find that whether wages are positively or negatively related to the level of monitoring depends critically on the shape of the worker's effort supply curve and, in particular, whether the elasticity of the worker's disutility of effort is increasing or decreasing in effort. The intuition behind Proposition 6 is simple and is found in a clear-cut interpretation of the efficiency wage schedule given by expression (16). The efficiency wage is greater the higher is the required non-shirking effort and the lower is the probability of shirker detection. If the shirkers response to non-shirker effort is sufficiently elastic it follows that the positive correlation between effort and wages can be broken. Thus, Proposition 6, viewed in isolation, raises concerns over the central efficiency wage tenet of a positive correlation between wages and effort.

The concern deepens when, in a similar manner, we draw conclusions regarding the level of non-shirker effort exertion and changes in unemployment insurance:

Proposition 7: A sufficient (but not necessary) condition for the negative supply-side correlation between unemployment insurance and (non-shirking) effort (i.e. $\partial e^/\partial b < 0$) is that the elasticity of the probability of detection with respect to the latter, η^* , is less than unity.*

Proof: The proof follows the proof of Proposition 6 closely and is therefore omitted.

QED.

Thus, and contrary to previous efficiency literature, we are no longer certain that higher unemployment insurance results in lower effort. As was the case for Proposition 6, this can be given some simple intuition through expression (16). The effort at a given efficiency wage should fall with the level of unemployment benefits, as long as the probability of detection of shirkers remain constant. However, the probability of detection does not remain unchanged, and herein lies the breakdown of the standard result.

We now turn to the firm's behaviour when it sets standards at such a level that no workers shirk in equilibrium. The analysis surrounding Propositions 6 and 7 is supply driven; rather than tying down a particular wage-effort combination, it investigated an incentive compatible locus of wage and effort combinations. To identify the equilibrium level of effort and the efficient wage from this locus, we turn to the demand side where the firm maximises profits subject to workers behaving according to their previously determined supply (i.e. pay-effort) schedule. Thus, armed with the knowledge of how workers respond in terms of effort to changes in pay, the firm will set the level of compensation that maximises profit. We now derive:

Proposition 8: *The firm will always choose an operational wage such that $\partial e^*/\partial w^* > 0$.*

Proof *From (2) and (27) it follows that:*

$$\begin{aligned} \frac{\partial \pi}{\partial e^*} &= \theta_i f'(e^*) - \frac{c'(e^*) - [c(e^*) - c(\bar{e})] \frac{\partial p}{\partial e^*}}{p^2} = 0 \\ \Rightarrow & \\ \theta_i f'(e^*) &= \frac{c'(e^*) - [c(e^*) - c(\bar{e})] \frac{\partial p}{\partial e^*} \frac{1}{p}}{p} \end{aligned} \tag{29}$$

Note that since the left hand side of (29) is positive the right hand side by deduction also has to be positive. From expression (28) the proposition follows.

QED.

Proposition 8 thus stands in contrast to the discussion following Proposition 6, which suggested that workers would, under certain circumstances, want to reduce their effort in response to an increase in wages. Indeed, it offers a resolution to the problematic result, contrary to the central premise of the efficiency wage literature, that higher wages might in some situations induce lower effort. Proposition 8 states that firms will always set wages such that the positive efficiency wage correlation between wages and effort holds. There is,

however, no internal conflict between the conditions that underpin Proposition 6 and Proposition 8. Instead it is merely a reflection of profit maximisation, for whilst Proposition 6 merely reflects supply responses, both demand and supply factors play a role in Proposition 8 thus assuring the best possible outcome for the firm in terms of profit. Note that given Proposition 8, and given the close relationship between Proposition 6 and Proposition 7, it must also be true that the firm operates in a region where an increase in unemployment insurance will induce a decline in effort.

This result is akin to the Stigler (1956) argument that a profit maximising monopolist should always operate at the elastic part of a monopolist's demand curve. Our paper argues similarly, by implication of Propositions 6 and 8, that a profit maximising firm will operate where the elasticity of detection is not too elastic. Given the efficiency wage schedule in expression (16), it will never be optimal to operate in the region where an increase in wage yields a lower effort, as this would reduce profits. Thus, given Proposition 6, the firm should operate where the probability of detection is not too elastic with respect to non-shirkers' effort. Proposition 8 shows formally that a profit maximising firm will always operate where an increase in wages yields larger levels of effort.

3. Final Comments

Our model as it stands illustrates a more nuanced picture regarding wages, effort and standards than previously acknowledged. Shirkers are no longer those workers who provide zero effort. They are instead those who neglect their duties by working less than required and who act rationally in so doing by trading off the cost of effort and the probability of detection. Thus, they work harder the higher the wage and the higher the standard set by the firm. By assuming continuous effort and endogenous detection, we identify conditions under which higher wages reduce effort whereas higher unemployment insurance increases effort, both of

which raise questions regarding the validity of the efficiency wage literature. We nevertheless offer a resolution to this set of two potentially disturbing results as our case is demonstrated to apply only to the supply side. For when we also take into account the demand side it becomes evident that the firm will always choose to operate in the region where workers respond to higher wages or lower unemployment insurance by increasing effort.

Monitoring technology has been central to large swathes of the efficiency wage literature. And whilst there are good and natural reasons for this, a departure from a focus on monitoring to one where observable output is used as a signal for effort has allowed us to construct a stochastic efficiency wage model within which we can investigate the largely neglected connections between standards and efficiency wages. The model we have proposed is one in which workers face idiosyncratic shocks to their output. As such, it is natural that the firm should use absolute performance criteria when considering firing. Were we to alter this assumption and consider the case where shocks instead are common to all workers, it may be more appropriate to use relative performance measures, with a relatively poor performance by a worker being used as the trigger mechanism resulting in a dismissal. Whilst such ‘avoid the drop’ tournaments are worthy of further investigation, they have remained unexplored here and are instead left for future research.

References

- Akerlof, G. A. and J. L. Yellen. (1990). ‘The Fair Wage–Effort Hypothesis and Unemployment.’ *Quarterly Journal of Economics*, 105, pp. 255–283.
- Akerlof, G. A. (1980). ‘A Theory of Social Custom, of which Unemployment May Be One Consequence.’ *Quarterly Journal of Economics* 94, no. 4 (1980): 749–75.
- Allgulin, M. and Ellingsen, T. (2002). ‘Monitoring and Pay.’ *Journal of Labor Economics*, 20, pp. 201–216.
- Blau, F.D. and Kahn, L.M. (1999). ‘Institutions and Laws in the Labor Market.’ In Ashenfelter, O. C. and Card, D. eds: *Handbook of Labour Economics* (North Holland).
- Calvo, G. A., and Wellisz, S. (1978). ‘Supervision, loss of control, and the optimum size of the firm.’ *Journal of Political Economy*, 86, 943–952.
- Clark, A. (2003). ‘Unemployment as a Social Norm: Psychological Evidence from Panel Data.’ *Journal of Labor Economics*, 21, pp. 323–51.
- Danthine, J. and A. Kurmann. (2006). ‘Efficiency Wages Revisited: The Internal Reference Perspective.’ *Economics Letters*, 90, pp. 278–284.
- Ewing, B. T. and Payne, J. E. (1999). ‘The Trade-off between Supervision and Wages: Evidence of Efficiency Wages from the NLSY.’ *Southern Economic Journal*, 66, pp. 424–432.

- Ghosh, S. and Waldman, M. (2010). ‘Standard Promotion Practices versus Up-or-Out Contracts.’ *Rand Journal of Economics*, 41, pp. 301-323.
- Gibbons, R. (1987). ‘Piece-Rate Incentives Schemes.’ *Journal of Labor Economics*, 5, pp. 413–429.
- Gibbons, R. and Waldman, M. (1999). ‘Careers in Organizations: Theory and Evidence.’ In O. Ashenfelter and D. Card, eds, *Handbook of Labor Economics*, Vol. 3. Amsterdam: North Holland.
- Goerke, L. (2001). ‘On the relationship between wages and monitoring in shirking models.’ *Metroeconomica*, 52, pp. 376–390
- Green, J. R. and Stokey, N. L. (1983). ‘A Comparison of Tournaments and Contracts.’ *Journal of Political Economy*, 91, pp. 349-364.
- Hahn, S. M. and Mayer, K. (2011). ‘Equilibrium Unemployment as a Discipline Device when finding Employment is Costly.’ *Canadian Journal of Economics*, 44, pp. 709-717.
- Healy, P. M. (1985). ‘The Effect of Bonus Schemes on Accounting Decisions.’ *Journal of Accounting & Economics*, 7, 85-112.
- Holmstrom, B. (1982). ‘Moral Hazard in Teams.’ *Bell Journal of Economics*, 13, pp. 324-340.
- Lazear, E. P. (1986). ‘Salaries and Piece Rates.’ *Journal of Business*, 59, pp. 405–431.
- Lazear, E. P. and Rosen, S. (1981) ‘Rank-Order Tournaments as Optimum Labor Contracts.’ *Journal of Political Economy*, 89, pp. 841-864.
- Malcomson, J. M. (1997). ‘Contracts, Hold-Up and Labor Markets.’ *Journal of Economic Literature*, 35, pp. 1916-1957.
- ‘Individual Employment Contracts.’ In O. Ashenfelter and D. Card, eds, *Handbook of Labor Economics*, Vol. 3. Amsterdam: North Holland.
- Mirrlees, J.A. (1974). ‘Notes on Welfare Economics, Information and Uncertainty.’ In Balch, M., McFadden, D. and Wu, S eds; *Essays on Economic Behavior under Uncertainty* (North Holland).
- Murphy, K. J. (1999). ‘Executive Compensation.’ in Ashenfelter, O. and Card, D. eds.: *Handbook of Labor Economics* (North Holland).
- Murphy, K. J. (2013). ‘Executive Compensation: Where We are and How We Got There.’ In Constantinides, G.M., M. Harris and Stulz, R.M. eds.: *Handbook of the Economics of Finance* (North Holland).
- Murphy, K. J., and Jensen, M.C. (2011). ‘CEO Bonus Plans and How to Fix Them.’ *Harvard Business School NOM Unit Working Paper 12-022*; Marshall School of Business Working Paper No. FBE 02-11.
- Prendergast, C. (1999). ‘The Provision of Incentives in Firms.’ *Journal of Economic Literature*, 37, pp. 7–63.
- Rasmusen, E. and Zenger, T. (1990). ‘Diseconomies of Scale in Employment Contracts.’ *Journal of Law, Economics and Organization*, 6(1), pp. 65-92.
- Shapiro, C. and J. E. Stiglitz. (1984). ‘Equilibrium Unemployment as a Worker Discipline Device.’ *American Economic Review*, 74, pp. 433-444.
- Solow, R. (1979). ‘Another Possible Source of Wage Stickiness.’ *Journal of Macroeconomics*, 1, pp. 79–82.
- Stigler, G. J. (1956). ‘The Statistics of Monopoly and Merger.’ *Journal of Political Economy*, 64, pp. 33-40.
- Stiglitz, J. E. (1985). ‘Equilibrium Wage Distributions.’ *Economic Journal*, 95, pp. 595-618.
- Strobl, E. and Walsh, F. (2007). ‘Estimating the Shirking Model with Variable Effort.’ *Labour Economics*, 14, pp. 623–637.
- Walsh, F. (1999). ‘A Multisector Model of Efficiency Wages.’ *Journal of Labor Economics*, 17, pp. 351-376.

Appendix

Proof of Proposition 3

The strict inequality $\bar{e} < e^*$ follows by definition. The strict inequality $\bar{e} > 0$ follows from first rewriting the expected utility from shirking as:

$$u(\bar{e}, e^*) = pb + (1-p)w - c(\bar{e}) = w + \left[\frac{\theta_L f(e^*)}{f(\bar{e})} - \theta_L \right] (b-w) - c(\bar{e}) \quad (\text{A1})$$

The optimal level of shirking effort is derived from the first- and second-order conditions:

$$\frac{\partial u(\bar{e}, e^*)}{\partial e} \equiv u_{\bar{e}}(\bar{e}, e^*) = \frac{\partial p}{\partial \bar{e}}(b-w) - c'(\bar{e}) = - \left[\frac{f'(\bar{e})}{f(\bar{e})^2} \right] \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] (b-w) - c'(\bar{e}) = 0 \quad (\text{A2})$$

$$\frac{\partial^2 u(\bar{e}, e^*)}{\partial \bar{e}^2} \equiv u_{\bar{e}\bar{e}}(\bar{e}, e^*) = \frac{\partial^2 p}{\partial \bar{e}^2}(b-w) - c''(\bar{e}) = - \left[\frac{f''(\bar{e})f(\bar{e}) - 2f'(\bar{e})^2}{f(\bar{e})^3} \right] \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] (b-w) - c''(\bar{e}) < 0 \quad (\text{A3})$$

Since $\lim_{e \rightarrow 0} f(e) = 0$ and $\lim_{e \rightarrow 0} f'(e) = \infty$, then $\lim_{e \rightarrow 0} [f'(e)/f(e)^2] = \infty$. Given that $\lim_{e \rightarrow 0} c'(e) = 0$ we have:

$$\lim_{\bar{e} \rightarrow 0} u_{\bar{e}}(\bar{e}, e^*) = - \left[\frac{f'(\bar{e})}{f(\bar{e})^2} \right] \left[\frac{\theta_L f(e^*)}{(\theta_H - \theta_L)} \right] (b-w) - c'(\bar{e}) > 0 \quad (\text{A4})$$

Supplying zero effort is therefore not optimal.

QED