

IZA DP No. 5619

Paid to Perform? Compensation Profiles under Pure Wage and Performance Related Pay Arrangements

John G. Sessions
John D. Skåtun

April 2011

Paid to Perform? Compensation Profiles under Pure Wage and Performance Related Pay Arrangements

John G. Sessions

*University of Bath
and IZA*

John D. Skåtun

University of Aberdeen Business School

Discussion Paper No. 5619

April 2011

IZA

P.O. Box 7240
53072 Bonn
Germany

Phone: +49-228-3894-0

Fax: +49-228-3894-180

E-mail: iza@iza.org

Any opinions expressed here are those of the author(s) and not those of IZA. Research published in this series may include views on policy, but the institute itself takes no institutional policy positions.

The Institute for the Study of Labor (IZA) in Bonn is a local and virtual international research center and a place of communication between science, politics and business. IZA is an independent nonprofit organization supported by Deutsche Post Foundation. The center is associated with the University of Bonn and offers a stimulating research environment through its international network, workshops and conferences, data service, project support, research visits and doctoral program. IZA engages in (i) original and internationally competitive research in all fields of labor economics, (ii) development of policy concepts, and (iii) dissemination of research results and concepts to the interested public.

IZA Discussion Papers often represent preliminary work and are circulated to encourage discussion. Citation of such a paper should account for its provisional character. A revised version may be available directly from the author.

ABSTRACT

Paid to Perform? Compensation Profiles under Pure Wage and Performance Related Pay Arrangements^{*}

Whilst existing efficiency wage literature assumes detection probabilities of shirkers are exogenous, this paper finds them positively and endogenously dependent on non-shirkers' effort. It shares the result with the endogenous monitoring models where, in some regions, workers reduce effort in response to higher wages, but differs in that firms never operate in those regions. The paper further provides theoretical reasons for the empirical regularity that increased usage of performance related pay (PRP) flattens the pay-tenure profile. Wages and effort increase over the lifecycle, both with and without PRP, but with late payments in PRP falling short of pure wage arrangements.

JEL Classification: J33, J41, J54

Keywords: monitoring, tenure, efficiency wages

Corresponding author:

John G. Sessions
Department of Economics
University of Bath
Bath
BA5 7AY
England
E-mail: j.g.sessions@bath.ac.uk

^{*} The normal disclaimer applies.

1. Introduction

Efficiency-wage theory predicts that firms can raise worker productivity by adopting a carrot and stick approach of paying a supra-competitive wage, devoting resources to monitoring, and dismissing any workers it detects as shirking [Shapiro and Stiglitz (1985)].¹ By extension, if employment contracts are multi-period then firms can raise worker productivity by *tilting* the remuneration package over time, paying workers less than the value of their marginal product when they are relatively short-tenured, and correspondingly more than the value of their marginal product when they are relatively long tenured. Such a pay profile provides workers with *ex post* rents that they will be reluctant to jeopardise. If reducing effort increases the probability of involuntary termination, then upward sloping pay profiles raise the cost of shirking and induce workers to raise effort.²

An alternative method of raising worker productivity is to divest a share of the firm into the hands of workers through collective (e.g. profit sharing, employee share ownership) and / or individual performance-related pay (PRP) schemes [see Weitzman and Kruse, (1990), Blinder (1990)]. Such schemes directly reduce the marginal benefit of shirking and therefore have implications for both the stationary and dynamic versions of the efficiency wage story. In terms of the latter, introspection would suggest - and empirical work has confirmed - that the greater the component of total pay that is derived from PRP, then the flatter is the slope of the pay-tenure profile *ceteris paribus* [Lazear and Moore (1984), Brown and Sessions (2006)].

In this paper we rationalise formally the relationship between PRP and the nature of the pay-tenure profile. Under not unreasonable assumptions, we find that effort and

1 The Shapiro-Stiglitz model regards worker effort as synonymous with productivity. This need not be the case. Other conduits through which efficiency wages might impact upon productivity include reduced turnover [Salop (1979), Stiglitz (1985)], adverse selection [Weiss (1980)] and worker morale [Akerlof (1982)].

2 Recent support for this prediction is offered by Adams and Heywood (2011) who find evidence from both US and Australian data that deferred compensation, whether in the form of steeper tenure-wage profiles or pensions, is associated with higher self-reported worker effort, with the increase in effort declining as the chance of job separation rises.

remuneration exhibit a greater tenure gradient under a wage only contract as compared to contracts where remuneration includes some element of PRP. With remuneration in pure wage contracts being higher later on in the firm-worker relationship, they are also compensatingly lower earlier on as compared to firm-worker contracts that include PRP. The results of the simple models presented below are therefore compatible with, and give theoretical support, to previous empirical findings.

A key objective of the efficiency wage literature that follows Shapiro and Stiglitz (1985) is to endogenize unemployment rates.³ Monitoring, in contrast, is usually regarded as exogenous, although its role remains critical since increases in monitoring typically raise effort and depress wages. Some recent work has begun to diverge from this tradition by endogenizing both unemployment and monitoring. Notable contributions here include Walsh (1999), (2001) and Strobl and Walsh (2007), both of whom find that increased monitoring yields ambiguous effects on wages.

There is an intriguing resemblance between the above contributions and our analysis. Our model differs from the above by focussing directly on the endogeneity of the probability of detection with respect to effort. We assume that firms are able to observe perfectly individual worker output, but not the effort and luck upon which output positively depends. Thus, when a worker's output is above some critical level, firms are unable to attribute this directly to worker effort or luck. Shirkers may, on the other hand, involuntarily expose their true nature when they are sufficiently unfortunate by producing a revealed level of output that is so low that it cannot conceivably be produced by a non-shirking worker. How lucky a shirker needs to be will therefore depend critically on the equilibrium effort level of non-shirkers. Were the non-shirkers to collectively exert a large equilibrium level of effort, the shirker would need to be extra lucky. There are, in other words, less states (of luck) within

³ A recent paper in this tradition is Basu and Felkey (2009) which finds multiple unemployment equilibria.

which to hide when the equilibrium level of effort is high. Thus, the *ex ante* probability of detecting a shirker is endogenized to depend positively on effort. We show that this may, in certain wage ranges, break the normal positive supply relationship between effort and wages. However, though the supply of effort may in some cases react perversely and drop in response to an increase in the wage, as is also the case in the endogenous monitoring literature, we nevertheless demonstrate that the more standard efficiency wage supposition of higher effort in response to higher wages is retained when both demand and supply side factors are taken into account. Indeed we find that it is never optimal for firms (on the demand side) to operate in regions of the wage (on the supply side) where effort falls in response to wage hikes.

The paper is set out as follows: Section 2 discusses the wage-seniority nexus whilst Section 3 sets out our modelling framework. Section 4 investigates how the supply of effort depends on pay under ‘pure wage’ and ‘PRP’ schemes, the latter being defined as schemes that consist of a pure wage term and a performance dependent term.⁴ Section 4 investigates the demand side decisions by the firm and Section 5 offers some concluding comments.

2. The Wage-Seniority Nexus

The correlation between seniority and pay is one of the most robust empirical findings in labour economics - for surveys of the theoretical and the empirical literature, see Hutchens (1989), Carmichael (1990), Polachek and Siebert (1992), and Lazear (2000). The source of the relationship, however, is somewhat ambiguous and in recent years two key theoretical explanations have emerged.

⁴ We use the term ‘contract’ in a loose sense and interchangeable with ‘payment arrangement’, where the firm does not pre-commit to a particular wage / remuneration profile. Any ensuing profiles of payments are rather a reflection of remuneration policies which are set on a single-period basis, but where period-one decisions are contingent on the realised pay of period-two.

The conventional explanation up until the 1980s for the relationship was that earnings reflected the acquisition of, and reward to, human capital. Workers became more productive, and hence better remunerated, over time because of investments in training. Such investments could be either specific or general; the former increased a worker's productivity in the worker's current firm, whilst the latter increased a worker's productivity both in the worker's current firm and in any future firm. Workers paid for general training, and subsidised specific training, by accepting early career (i.e. training) wages below the value of their marginal product to the firm. Later career (i.e. trained) wages reflected the increase in worker productivity; fully in the case of general training and partially in the case of specific training. Since specific training is only of value within a worker's current firm, it is optimal for workers to neither pay the full cost nor reap the full benefit of such training - to do otherwise might tempt the firm into making redundancies in an attempt to replace trained with untrained workers. By paying a trained wage at a rate below the value of a trained workers' marginal product, firms are dissuaded from laying-off trained workers and, accordingly, workers are persuaded to participate in specific training programs. In either case an upward sloping wage profile emerges; wages increase with seniority because productivity increase with seniority [Mincer (1958), Becker (1962), Ben-Porath (1967)].

The human capital explanation was challenged in a series of papers by Lazear (1979, 1981, 1983) and Medoff and Abraham (1980, 1981). Lazear observed that mandatory retirement and actuarially unfair pension schemes that encourage early retirement were incompatible with human capital theory. Why would firms establish human resource policies whereby an employee was paid, and thus evidently valued, today but then either forced or induced to quit tomorrow? Such policies contradict the human capital thesis that senior workers are paid no more than their marginal product, particularly when wages can be adjusted downwards if productivity declines with age.

Medoff and Abraham (1980, 1981) highlighted a related conundrum in their analysis of data on pay and supervisor performance ratings. They found that although relative performance ratings within a particular job grade did not increase with experience in the job grade, relative pay did. Again, such a finding is incompatible with the human capital position that earnings increase with seniority because productivity increases with seniority

Several models of wage setting are able to explain the apparent failings of the human capital model. Freeman (1977) and Harris and Holmström (1982) develop models in which risk averse workers prefer upward sloping wage profiles because they offer insurance against the risk that the workers' future productivity is lower than anticipated. Another possibility is that workers prefer rising consumption profiles over their life cycle but find voluntary saving difficult. Upward sloping earnings profiles are therefore desirable because they represent a mechanism for forced-saving [Loewenstein and Sicherman (1991), Frank and Hutchens (1993), Neumark (1995)]. And models of job search generally predict that more time in the labor market increases the chance of finding a better match and thus tends to be associated with higher earnings [Burdett (1978), Ruhm (1991), Jacobson and LaLonde (1993), Manning (2000)].

Perhaps the most persuasive explanation is the agency approach developed by Lazear (1979, 1981, 1983). Lazear reconciled the various phenomena by focussing on contracts that discourage employee shirking and other malfeasance over an employee's life cycle, especially in situations where monitoring worker effort is problematic. The basic idea is that workers and firms enter into contracts, implicit or explicit, whereby workers are paid less than the value of their marginal product when they are in the early years of their job tenure, and correspondingly more than the value of their marginal product when they are in their latter years. By back-loading compensation in this way, workers are provided with *ex post* rents that they are reluctant to lose. If reducing effort increases the probability of involuntary

termination, then upward sloping wage profiles raise the cost of shirking and encourage workers to raise effort.

Lazear's explanation cuts the link between productivity and pay; wages grow with seniority irrespective of how production relates to seniority. And whilst it makes sense for the firm to pay wages in excess of the value of a worker's marginal product for a period of time, it would not be sensible for the firm to do this indefinitely. There will come a point when the present discounted value of the worker's marginal product equals the present discounted value of his remuneration package. This would imply, from the firm's perspective, an optimal retirement date and hence the need for policies to force or encourage the worker's retirement.

The question as to whether it is human capital or agency considerations that drive the wage profile is not entirely academic. If human capital explanations are found wanting and agency considerations dominate, then issues arise concerning the credibility of long-term employment contracts. Firms would prefer to shed senior workers who are more expensive but perhaps not more productive. This could lead to time-consistency problems, with some firms finding it difficult to attract younger applicants because of their inability to guarantee long-term employment. If, on the other hand, the slope is primarily a reflection of human capital considerations, then such incentive-compatibility problems will not arise - older workers will be more productive *ceteris paribus*. In this case, the wage profile provides some indication of the return to investments in on-the-job training and education.⁵

Several studies have attempted to discriminate empirically between the two explanations. Hutchens (1987) focuses on the implicit trade-off between the use of deferred payment contracts and the difficulty of monitoring and finds that Lazear-type characteristics (i.e. wage profiles, mandatory retirement, pension schemes, long job tenures) tend not to be associated with jobs that are conducive to monitoring.

⁵ The profile may also affect quitting behaviour. More experienced 'generally' trained workers will have more flexibility in the labour market than 'specifically' trained workers. But both types may have more options than senior workers whose rents primarily reflect agency considerations.

Lazear and Moore (1984) address the issue by considering the empirical evidence regarding the relative ‘flatness’ of self-employed workers’ wage profiles [Wolpin (1977), Fuchs (1981)]. Such a finding is puzzling since investments in physical capital would tend to depress observed wages for the early career self-employed, whilst subsequent returns to those investments would tend to raise observed wages. Both factors imply that, other things equal, the wage profiles of self-employed workers will be steeper than those of wage and salary workers.

Lazear and Moore (1984) rationalize the finding by highlighting the duality of principal and ownership intrinsic to self-employment. Observed wage profiles, they argue, are a reflection of the disharmony of interests prevalent in the employment relation, a dissonance that is, by definition, absent from self-employment. By steepening the wage profile, employers are able to induce their employees to work harder, therefore raising the present value of the latter’s lifetime earnings. The self-employed require no such internal incentive mechanism and thus may be used as a control group to test the theoretical prior that the profile is determined primarily by agency as opposed to human capital considerations.

Brown and Sessions (2006) generalise Lazear and Moore’s approach by comparing the earnings profiles of self-employed workers, wage / salary workers, and workers employed under PRP schemes. If agency considerations are important in driving the earnings profile, and if PRP workers face an intermediate degree of agency as compared to wage / salary and self-employed workers, then the earnings of PRP workers should increase at an intermediate rate with tenure as compared to these other types of workers.

Both studies find convincing empirical support – Lazear and Moore from U.S. data, Brown and Sessions from British data - for the argument that it is agency considerations that drive the pay-tenure profile. In what follows, we endeavour to formally rationalise these findings.

3. The Modelling Framework

Our aim in the following two period model is to illustrate the relationship between, agency, worker ‘equity’ (i.e. the extent of any PRP) and the nature of the earnings profile. For ease of analytical exposition, we therefore abstract from considerations regarding human capital and focus instead on the supply and demand aspects of cost-minimising contracts offered by a firm in the presence of asymmetric information regarding worker effort. Two regimes are considered, a ‘pure wage’ regime, where firms can only use (fixed) wages to compensate workers, and a PRP regime, where remuneration in each period consists of two elements; a fixed wage and an additional component related to individual worker performance. This section outlines modelling aspects common to both regimes.

We assume that workers are homogenous, risk neutral and endowed with a working life of two periods. Their separable periodic utility function, u_t , is given by $u_t = m_t - g(e_t)$, $t = 1, 2$, where m_t denotes income in period t , e_t denotes the worker’s supply of effort in period t , and $g(\cdot)$ is a continuous cost function with $g(0) = 0$ such that no cost is incurred at zero effort. The usual cost properties apply; of being increasing in effort, $dg(e_t)/de_t \equiv g'(e_t) > 0$, and convex in effort, $d^2g(e_t)/de_t^2 \equiv g''(e_t) > 0$. We assume that employed workers make an all or nothing choice as regards the provision of effort to their employer: That is they either do not shirk but supply effort, $e_t > 0$, or they shirk and provide zero effort, $e_t = 0$. The downside to a worker of the latter option is the chance of being detected shirking and fired, where the probability $p_t = p \in (0,1)$ of a ‘shirker’ being detected is determined endogenously. We assume that detection of shirking implies instantaneous dismissal and (permanent) unemployment with a per-period utility $b_t = b > 0$.⁶

⁶ Whilst we assume that detected shirkers are fired and forced into *permanent* unemployment, this is an expository device only and imposed with no loss of generality. Indeed, allowing a more realistic scenario whereby detected shirkers receive

Each worker has a stochastic production/output/revenue function, $y_{it} = \theta_{it}f(e_t)$, denoting the revenue associated with her/his job in state i at time t . We assume here that $df(e_t)/de_t \equiv f'(e_t) > 0$, $d^2f(e_t)/de_t^2 \equiv f''(e_t) < 0$ and $f(0) > 0$, such that output increases at a diminishing rate with effort and is positive even if the worker shirks. The shift-parameter, θ_{it} , represents a random shock to productivity in state i at time t and is uniformly distributed between two values: a lower value θ_L , and an upper value $\theta_H > \theta_L$. For an individual worker, θ_i reflects the relative misfortune (when it is low) or luck (when it is high). Whilst the firm is able to observe worker revenue, it is unable to observe either worker effort, e_t , or 'luck', θ_{it} . Nonetheless, it seems reasonable to assume that effort can in some instances be deduced. To reflect this, we define a critical realisation of the random shock, θ_{ct} , at or above which it is impossible at time t to deduce whether the worker is a shirker or a non-shirker. Formally, we assume that: $\theta_{it}f(e_t) \geq \theta_L f(e_t) = \theta_{ct}f(0)$, $\forall \theta_i \in [\theta_L, \theta_H]$. Thus, the critical state satisfies $\theta_{ct} \equiv \theta_{ct}(e_t) = \theta_L f(e_t)/f(0)$, which by necessity is increasing in the equilibrium level of effort since $d\theta_{ct}/de_t \equiv \theta'_{ct}(e) = \theta_L f'(e_t)/f(0) > 0$. Intuitively, a higher equilibrium level of effort on the part of non-shirkers raises the acceptable bar of revenue performance, resulting in more states in which shirking behaviour is identifiable. Thus, those worker who, despite the higher equilibrium level of non-shirking effort, continues to shirk 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 Figures 1 and Figure 2. The two upward sloping lines in Figure 1 depict the revenues generated by a shirking worker, $y_{it} = \theta_{it}f(0)$, and a non-shirking worker, $y_{it} = \theta_{it}f(e_t)$. Since the firm is only able to observe revenue, but not its

unemployment benefits in period one and then have a chance of obtaining a (single period) employment contract in period two would not change our qualitative results.

constituent elements (i.e. effort and luck), it is unable to distinguish between a non-shirker whose productivity realisation is $\theta_{ct}(e_t)$ and a shirker whose productivity realisation is at the lower bound θ_L . More generally, the firm is unable to detect shirking at any productivity realisation $\theta_{it} \geq \theta_{ct}(e_t)$. Shirking is, however, detectable at productivity realisations $\theta_{it} < \theta_{ct}(e_t)$ since the revenue from shirking here falls short of the lowest output possible for a non-shirker.

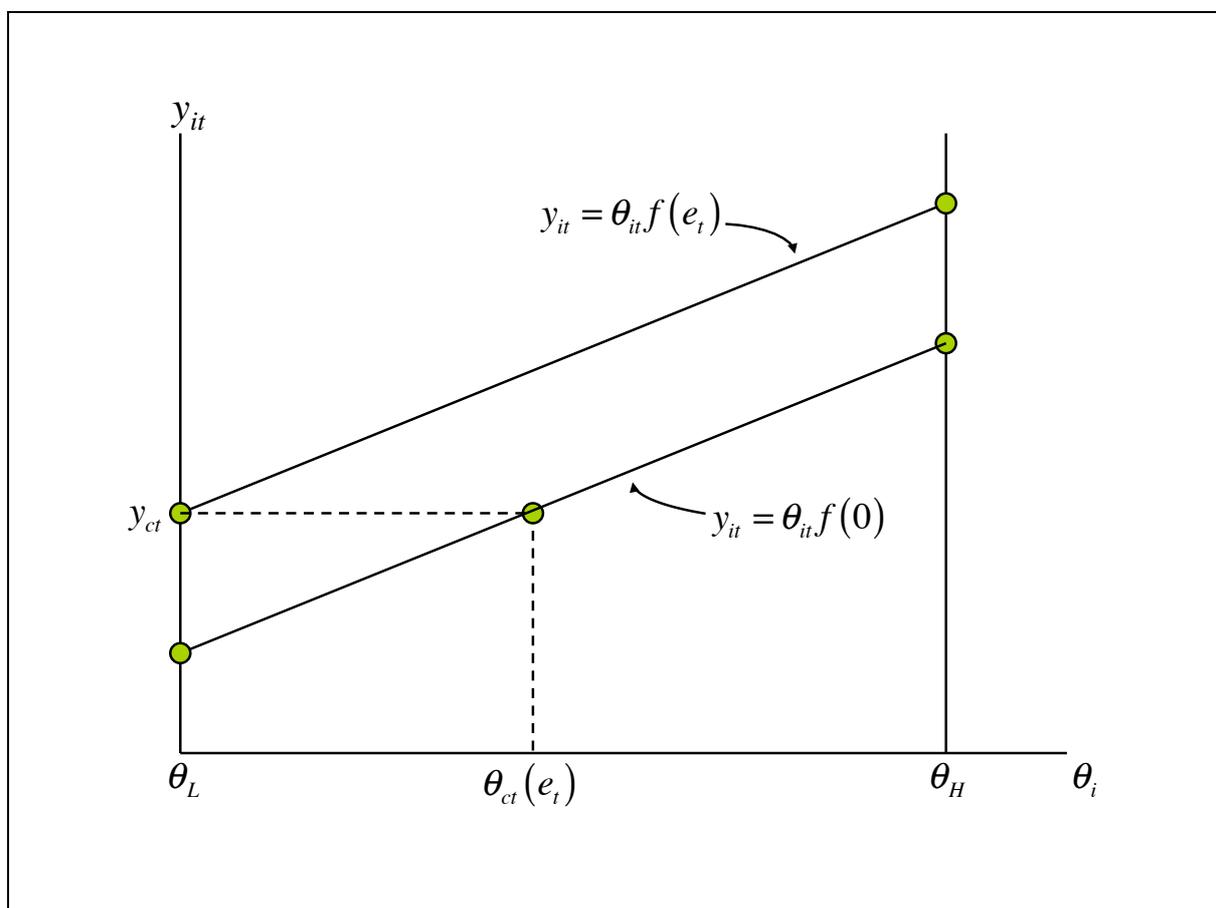


Figure 1: Critical 'Luck'

An increase in the equilibrium non-shirking level of effort from e_t^0 to $e_t^1 > e_t^0$ in Figure 2 increases the critical shift parameter from $\theta_{ct}(e_t^0)$ to $\theta_{ct}(e_t^1) > \theta_{ct}(e_t^0)$. Thus, as the lowest possible output for non-shirkers increases, shirkers need to be even luckier in order to avoid detection. Assuming only shirking workers are fired, the probability, p_t , of a shirker being

fired is for a given level of effort given by $p_t \equiv p_t[\theta_{ct}(e_t)] = [\theta_{ct}(e_t) - \theta_L] / (\theta_H - \theta_L)$. In contrast to the conventional efficiency wage story, this probability is determined endogenously by the equilibrium level of effort. Indeed, we derive:

Proposition 1: *The probability of detection if shirking depends positively on the equilibrium effort levels of non-shirkers.*

Proof: *This follows simply from $dp_t/de_t = \theta'_{ct}(e_t) / (\theta_H - \theta_L) > 0$.*

QED.

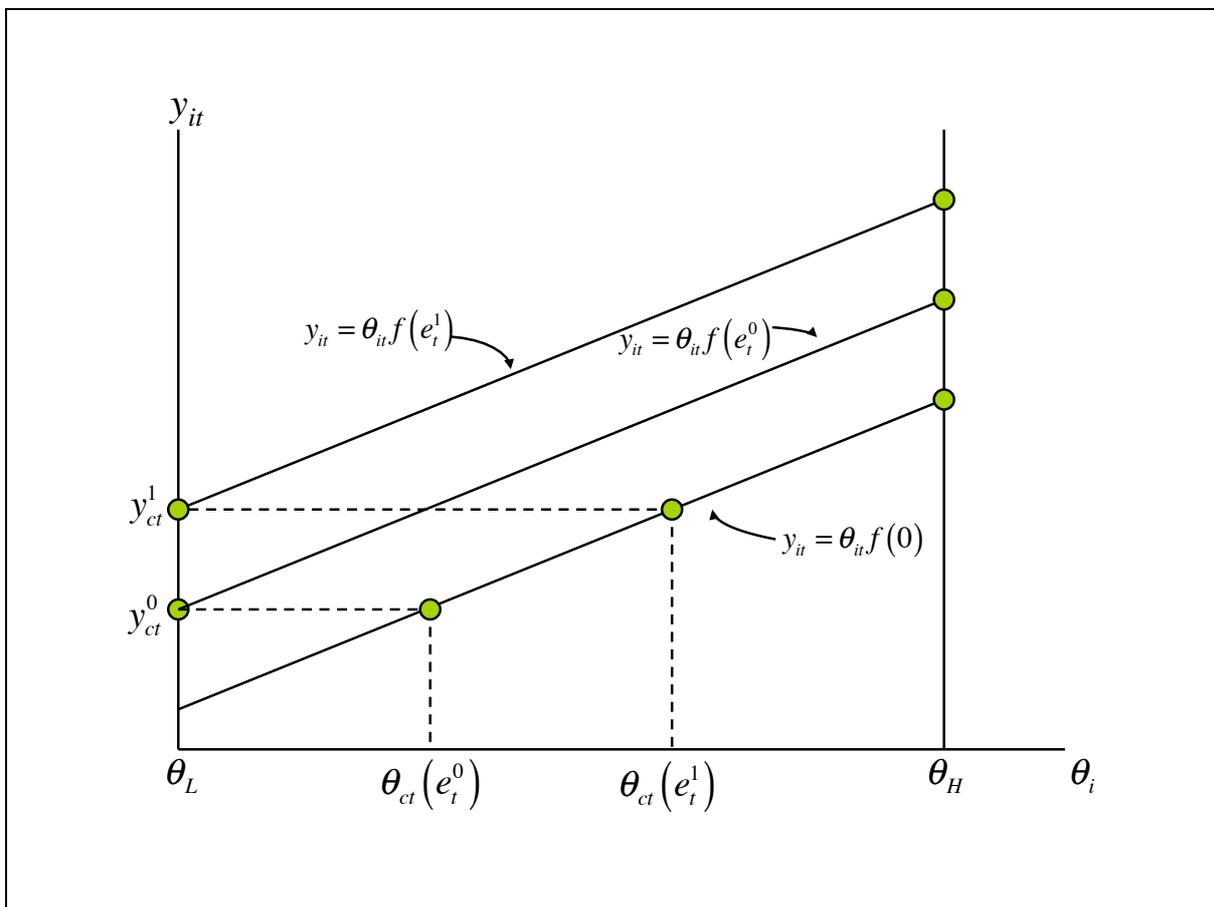


Figure 2: Critical Luck and Effort.

As Proposition 1 states and Figure 2 illustrates, the probability of detecting (and thus dismissing) a shirker will increase 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 shirkers who stand out more when average effort levels within a firm rise. Proposition 1 is thus in sharp contrast to

previous literature where effort and detection probabilities are unrelated and provides a possible route through which probabilities of a shirker being found out are endogenized to depend positively on effort levels.⁷

To determine optimal remuneration and its relationship with effort, in both the pure wage and the PRP case, we investigate the supply and demand side responses of effort to changes in remuneration. Starting with supply issues we first determine the incentive compatible remuneration-effort locus. Assuming that the firms can anticipate this supply side response we then use this same supply locus to determine demand side behaviour that is reflected in the remuneration that firms will optimally set. The overall efficiency remuneration/wage therefore incorporates both supply and demand side issues.

4. The Supply Side: Incentive Compatible Effort and Pay

This section determines the supply relationship between worker effort and remuneration under both wage-only and PRP regimes. There are two ways to view this analysis: It may either be looked upon as a derivation of the incentive inducing effort levels at a given level of pay; or alternatively as an investigation into the lowest remuneration required to induce a given level of effort, though it effectively makes no difference since both interpretations yield the supply locus of remuneration and effort. In order to distinguish between the two payment regimes, we denote the time dependent level of effort in the wage only (i.e. salary only) setting as \tilde{e}_t , whilst it in the presence of PRP is denoted as \hat{e}_t . We now turn to investigate the supply side of the two payment schemes in turn.

4.1 Incentive Compatible Remuneration under a Wage-Only Regime

We investigate first the supply locus of remuneration and effort under the wage only setting.

⁷ 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 outlined in Akerlof and Yellen (1990) and more recently in Danthine and Kurmann (2009) where the central theme is the relative wage within the firm as opposed to an external reference wage.

Consider the one period (spot wage) case. For a given level of effort, \tilde{e}_t , there exists a wage, \tilde{w}_t , at or above which the worker will always supply (at least) \tilde{e}_t and below which the worker will always shirk. Paying the worker \tilde{w}_t will, in other words, be sufficient to induce the worker to supply effort \tilde{e}_t . Note that in order to induce the worker to supply more effort than \tilde{e}_t , the wage would have to rise beyond \tilde{w}_t . Such supply considerations can be summarised by the incentive compatible ‘non-shirking constraint’ (NSC), which specifies the lowest wage required for the worker to supply a given effort level. Alternatively, it specifies the maximum effort supplied at a given wage. In the spot market one-period t case, the NSC in the wage only contract, denoted NSC_s^w , merely states that the effort exerting utility, $\tilde{w}_t - g(\tilde{e}_t)$ must be at least as high as the expected utility, $\tilde{p}_t b + (1 - \tilde{p}_t)\tilde{w}_t$, of shirking, where this is the weighted average of unemployment payoff b if detected with probability \tilde{p}_t and ‘getting away with it’ with probability $(1 - \tilde{p}_t)$ whilst collecting the wage \tilde{w}_t :

$$NSC_s^w : \quad \tilde{w}_t - g(\tilde{e}_t) \geq \tilde{p}_t b + (1 - \tilde{p}_t)\tilde{w}_t \quad (1)$$

Satisfaction of NSC_s^w implies an incentive compatible (i.e. efficiency) wage schedule of:

$$\tilde{w}_t = b + \frac{g(\tilde{e}_t)}{\tilde{p}_t} \quad (2)$$

This gives us the required wage at any given level of effort such that workers are just indifferent between shirking and not shirking. Note here that the incentive compatible wage level is a function of three elements: It is increasing in terms of the worker’s outside unemployment opportunity, b , since the firm will have to pay the worker more to induce effort when alternative employment prospects are good. Second, the wage is high when the cost of effort, $g(\tilde{e}_t)$, is high since the firm will need to pay the worker more to induce effort,

the more costly it is for the worker to do so. And third, the wage is lower the higher the probability of detecting a shirker, where fear of detection drives the worker to exert higher effort. Higher detection probabilities thus shade the necessary effort-inducing wage the firm is compelled to offer. Note that since both the cost of effort and the probability of detection are positively correlated with effort, but oppositely correlated to the wage rate, a new complexity has arisen whereby the relationship between the incentive compatible wage and the effort level is not unambiguously positive.

Proposition 2: *The positive correlation on the supply side between the (efficiency) wage and the effort level (i.e. $\partial \tilde{e}_i / \partial \tilde{w}_i > 0$) holds as long as the probability elasticity of effort is no greater than unity.*

Proof: *See Appendix.*

Proposition 2 illustrates that we are unable to rule out a fissure in the positive efficiency wage link between remuneration and the optimal supply of 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 (disutility of effort), are we able to retain the intuitively attractive positive correlation between effort supplied and efficiency wages prevalent in the literature. 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.

Now consider the specification of a two-period 'lifetime' wage scheme $(\tilde{w}_1, \tilde{w}_2)$. At the start of the second period, the firm and the worker are locked together in an employment relationship they know will only last for one period and as such they face the same effort elicitation considerations as firms and workers in a single-period 'spot market'. Thus, the

wage-effort schedule in period two is naturally and commonly known to be given by expression (2) such that $\tilde{w}_2 = b + g(\tilde{e}_2)/\tilde{p}_2$. It is also common knowledge that the firm in period two will take this schedule and choose the efficiency wage, later to be determined, which maximises its profit. Hence both firms and workers in period one are able to correctly anticipate the second period efficiency wage, such that the second period equilibrium outcome has first period effects. That is to say, the worker's first period effort compatible inducing wage will now be partially dependent on future wages. If, for example, the future wage is relatively high, then the worker will at any given level of period one wages have more to lose if detected shirking and fired in period one. This will in turn lower the necessary period one wage required to elicit a particular level of effort.

Formally, the *lifetime* period one NSC in the wage only contract with no discounting, denoted NSC_L^w , is given by:

$$NSC_L^w : \quad \tilde{w}_1 + \tilde{w}_2 - g(\tilde{e}_1) - g(\tilde{e}_2) \geq \tilde{p}_1(2b) + (1 - \tilde{p}_1)[\tilde{w}_1 + \tilde{w}_2 - g(\tilde{e}_2)] \quad (3)$$

Undetected shirkers enjoy utility of \tilde{w}_1 now and $\tilde{w}_2 - g(\tilde{e}_2)$ tomorrow, where \tilde{w}_2 is the incentive compatible wage schedule that ensures workers do not shirk in period two but instead exert the required second period effort. Detected shirkers receive b in both periods.⁸ Satisfaction of (NSC_L^w) together with the period-two version of (2) implies the period one incentive compatible wage-effort schedule:

$$\tilde{w}_1 = b + g(\tilde{e}_2) + \frac{g(\tilde{e}_1)}{\tilde{p}_1} - \frac{g(\tilde{e}_2)}{\tilde{p}_2} \quad (4)$$

⁸ Note the assumption that detected shirkers are fired and forced into *permanent* unemployment. This is an expository device. Allowing a more realistic scenario whereby detected shirkers receive unemployment benefits in period one and then have a chance of obtaining a (single period) employment contract in period two would not change our qualitative results.

Though we have not yet explicitly determined neither the profit maximising efficient wages in period one and period two, nor their associated effort levels, these variables will nevertheless have to comply with the supply conditions (2) and (4) respectively. Thus, any resulting wage profile will have to satisfy the following:

$$\Delta\tilde{w} \equiv \tilde{w}_2 - \tilde{w}_1 = \left(\frac{2 - \tilde{p}_2}{\tilde{p}_2} \right) g(\tilde{e}_2) - \left(\frac{1}{\tilde{p}_1} \right) g(\tilde{e}_1) \quad (5)$$

While we will infer more about the nature of the earnings profile under pure wage contracts later, we are presently unable to deduce whether the profile rises or falls with tenure. We can, however, still note the special case where $g(\tilde{e}) = g(\tilde{e}_1) = g(\tilde{e}_2)$ and $\tilde{p} = \tilde{p}_1 = \tilde{p}_2$ in which (5) reduces to $\Delta\tilde{w} = [(1 - \tilde{p})/\tilde{p}]g(\tilde{e}) > 0$. Intuitively, workers acquire rents on account of the firm's inability to perfectly monitor. The firm, however, can reduce these rents by offering lifetime contracts, *irrespective of human capital considerations*, that induce workers to queue up to gain access to the second period wage which exceeds their reservation utility.

4.2 Incentive Compatible Remuneration under PRP

Consider now a performance related pay (PRP) scheme in which total *actual* remuneration, \hat{w}_2^a , comprises both a 'fixed' wage component, \bar{w}_2 , and a variable 'performance related' output payment, with λ reflecting the relative weight attributed to the performance element of the payment such that $\hat{w}_2^a = (1 - \lambda)\bar{w}_2 + \lambda\theta_{it}f(\hat{e}_2)$.⁹ Workers are unable to observe the state of nature, θ_{it} , before they have exerted effort and so their decision regarding shirking will depend on total *expected* remuneration, \hat{w}_2 , which is given by:

$$\hat{w}_2 = (1 - \lambda)\bar{w}_2 + \lambda E\{\theta_{it}f(\hat{e}_2)\}, \text{ where } E\{\theta_{it}f(\hat{e}_2)\} \text{ denotes the expected second period}$$

⁹ We assume in what follows that the extent PRP, as measured by λ , is exogenous. This is obviously a simplistic assumption and a more complete exposition would seek to explain the distribution of different contractual arrangements.

output of the worker. The spot market/ period two NSC in PRP, denoted NSC_s^{pp} , takes the form:

$$NSC_s^{pp} : (1-\lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(\hat{e}_2)\} - g(\hat{e}_2) \geq \hat{p}_2 b + (1-\hat{p}_2) \left[(1-\lambda)\bar{w}_2 + \lambda E\{\theta_{i2}f(0)\} \right] \quad (6)$$

Here, the fixed wage component of total period two remuneration is given by \bar{w}_2 and the expected output for period two non-shirkers and (undetected) shirkers is given by $E\{\theta_{i2}f(\hat{e}_2)\} = [(\theta_H + \theta_L)/2]f(\hat{e}_2)$ and $E\{\theta_{i2}f(0)\} = [(\theta_H + \theta_{c2})/2]f(0)$ respectively, where $\theta_{c2} \equiv \theta_{c2}(e_2) = \theta_L f(e_2)/f(0)$ is the critical period-two realisation of luck determined previously. It is apparent from the above that the expected output of a non-shirker is increasing in effort. That the expected output of the undetected shirker is also positively correlated with effort, as can be shown, is perhaps less obvious.¹⁰ Though this conclusion may be surprising initially, it is nevertheless rooted in some fairly straightforward intuition: As the equilibrium level of effort increases, it becomes increasingly less likely that a shirker will go undetected. As effort increases amongst the non-shirkers therefore, the shirker rides his luck to an even greater extent and must be even more fortunate in the realisation of her/his θ , for there are less states of nature within which the lazy worker is able to hide when equilibrium effort amongst the thrifty is high. With the output of a shirker being determined jointly by the zero effort decision and the random variable, θ_{ii} , it follows that a higher realisation of the latter will increase the production of the shirker and so the relationship follows.

The NSC expression (NSC_s^{pp}) reduces to solve for the second period wage-effort schedule:

¹⁰ This follows directly from $\partial E_2\{\theta_{i2}f(0)\}/\partial \hat{e}_2 = (\partial \theta_{c2}/\partial \hat{e}_2)[f(0)/2] = [(\theta_H - \theta_L)/2](d\hat{p}_2/d\hat{e}_2)f(0) > 0$, since $d\hat{p}_2/d\hat{e}_2 = \theta_{c'}(\hat{e}_2)/(\theta_H - \theta_L) > 0$ as shown previously.

$$\bar{w}_2 = \frac{\hat{p}_2 b + g(\hat{e}_2) - \lambda \left[E\{\theta_{i2} f(\hat{e}_2)\} - (1 - \hat{p}_2) E\{\theta_{i2} f(0)\} \right]}{(1 - \lambda) \hat{p}_2} \quad (7)$$

Note, however, that it is not the wage component per se that determines whether the worker supplies the given effort. Rather, it is the total level of compensation, which consists of both a wage and a performance related element. With the use of (7) we can determine the period-two total (expected) remuneration, \hat{w}_2 -effort schedule:

$$\hat{w}_2 = (1 - \lambda) \bar{w}_2 + \lambda E\{\theta_{i2} f(\hat{e}_2)\} = b + \frac{g(\hat{e}_2)}{\hat{p}_2} - \left(\frac{1 - \hat{p}_2}{\hat{p}_2} \right) \lambda \Delta E_2 \quad (8)$$

where $\Delta E_2 = E\{\theta_{i2} f(\hat{e}_2)\} - E\{\theta_{i2} f(0)\}$. To assure worker participation in the second period employment relationship, we assume that the utility derived within the firm is at least equal to the utility of being unemployed *vis.* $\hat{w}_2 - g(\hat{e}_2) \geq b$. From this and (8) we deduce that:

$$\hat{w}_2 - b - g(\hat{e}_2) = \left(\frac{1 - \hat{p}_2}{\hat{p}_2} \right) \left[g(\hat{e}_2) - \lambda \Delta E_2 \right] \geq 0 \quad (9)$$

The period-two incentive compatible pay, \hat{w}_2 , is anticipated in period one and will therefore enter the workers' period-one lifetime NSC, denoted NSC_L^{pp} , which can be expressed as:

$$NSC_L^{pp} : \quad \begin{aligned} & (1 - \lambda) \bar{w}_1 + \lambda E\{\theta_{i1} f(\hat{e}_1)\} + \hat{w}_2 - g(\hat{e}_1) - g(\hat{e}_2) \\ & \geq \\ & \hat{p}_1 2b + (1 - \hat{p}_1) \left\{ (1 - \lambda) \bar{w}_1 + \lambda E\{\theta_{i1} f(0)\} + \hat{w}_2 - g(\hat{e}_2) \right\} \end{aligned} \quad (10)$$

Solving for the equilibrium period one fixed wage element, \bar{w}_1 , yields the following wage-effort schedule:

$$\bar{w}_1 = \frac{2\hat{p}_1 b - \hat{p}_1 \hat{w}_2 + g(\hat{e}_1) + \hat{p}_1 g(\hat{e}_2) - \lambda \left[E\{\theta_{i1} f(\hat{e}_1)\} - (1 - \hat{p}_1) E\{\theta_{i1} f(0)\} \right]}{(1 - \lambda) \hat{p}_1} \quad (11)$$

Expected period-one total pay, \hat{w}_1 , for a non-shirker is thus given by:

$$\begin{aligned}\hat{w}_1 &= (1-\lambda)\bar{w}_1 + \lambda E\{\theta_{il}f(\hat{e}_1)\} \\ \Rightarrow \\ \hat{w}_1 &= b + g(\hat{e}_2) + \left[\frac{g(\hat{e}_1)}{\hat{p}_1} - \frac{g(\hat{e}_2)}{\hat{p}_2} \right] + \left[\left(\frac{1-\hat{p}_2}{\hat{p}_2} \right) \lambda \Delta E_2 - \left(\frac{1-\hat{p}_1}{\hat{p}_1} \right) \lambda \Delta E_1 \right]\end{aligned}\quad (12)$$

where $\Delta E_1 = E\{\theta_{il}f(\hat{e}_1)\} - E\{\theta_{il}f(0)\}$. Note the specific case where $g(\hat{e}_1) = g(\hat{e}_2) = g(\hat{e})$, which implies $\hat{p}_1 = \hat{p}_2 = \hat{p}$ and $\Delta E_1 = \Delta E_2$ such that $\hat{w}_1 = b + g(\hat{e})$. In words, when effort (and thus the cost of effort) do not change across the two time periods there is no discernible difference between period one remuneration across the two payment regimes, although second period differences in remuneration persist.

Though we have not yet defined the efficiency remuneration in each period that maximises the profit of the firm, the payment profile must, as was the case for the pure wage arrangement, satisfy the supply conditions as given above. From (8) and (12) we find that the earnings (total remuneration) profile with variable worker equity must comply with:

$$\Delta \hat{w} \equiv \hat{w}_2 - \hat{w}_1 = \left[\left(\frac{2-\hat{p}_2}{\hat{p}_2} \right) g(\hat{e}_2) - \left(\frac{1}{\hat{p}_1} \right) g(\hat{e}_1) \right] - \left\{ \left[\frac{2\lambda(1-\hat{p}_2)\Delta E_2}{\hat{p}_2} \right] - \left[\frac{\lambda(1-\hat{p}_1)\Delta E_1}{\hat{p}_1} \right] \right\} \quad (13)$$

4.3 *Supply Side Comparisons between Regimes*

We observe by comparing (13) with (5), that it is theoretically possible for the gradient of remuneration under PRP to be less than the remuneration gradient under pure wage contracts. Note however that the opposite is also possible and that which of the two regimes has the steepest remuneration profile will be dependent on the level of effort. Consider the special case when effort levels are invariant across time and payment schedules implying the fixed wage profile is given by $\Delta \tilde{w} = [(1-\tilde{p})/\tilde{p}]g(\tilde{e}) > 0$ as previously shown, whereas in the PRP

case we have $g(\hat{e}_1) = g(\hat{e}_2) = g(\hat{e})$, which implies $\hat{p}_1 = \hat{p}_2 = \hat{p}$ and $\Delta E_1 = \Delta E_2 \equiv \Delta E$ so that the PRP pay profile can be expressed by $\Delta \hat{w} = [(1 - \hat{p})/\hat{p}][g(\hat{e}) - \lambda \Delta E] \geq 0$. Observe from these two expressions that the wage profile in the wage only contract is strictly upward sloping and the remuneration profile under PRP is non-negative by implication of (9). Note also that the profile under PRP is flatter than under a wage only contract.

More generally, when effort levels are not time invariant, then we also note from a comparison of (5) and (13) that the shape of the earnings profile under pure salary and PRP schemes coalesce when either the extent of PRP or the impact of shirking on the PRP component of remuneration approach zero (i.e. as $\lambda \rightarrow 0$ or $\Delta E_1 \rightarrow \Delta E_2 \rightarrow \Delta E \rightarrow 0$).

5. Demand Side Issues and Efficiency Pay

The previous section utilised the non-shirking condition to define the supply of effort and its relation to pay. With a whole locus of incentive compatible effort and remuneration combinations, we are however only somewhat closer to determining what the actual levels of compensation and effort should be. In order to determine these equilibrium values of earning and effort we now introduce the demand side, where the firms set an ‘efficiency compensation’¹¹ to maximise its profits subject to the workers behaving according to their previous determined supply pay-effort schedule. The firm is, in other words, armed with the knowledge of how the workers in terms of effort respond to changes in pay. And it is this knowledge the firm uses to its own advantage when setting the compensation that induces the equilibrium effort that maximises the firm’s profit. For clarity of exposition, we distinguish individual revenue from the (expected) aggregate revenue, $Y_t = F(e_t N_t)$, of employing N_t identical workers in period t .

¹¹ We will use the term efficiency compensation/pay/earnings/remuneration rather than efficiency wage when compensation includes more than a pure wage element.

We will in terms of modelling critically assume that the firm cannot commit to period two remuneration in period one.¹² Instead it is left to choose the period specific control variables; wages and employment that maximise the period specific profit; $\pi_t = F(e_t N_t) - w_t N_t$, where $t = 1, 2$. It is well known, following Solow (1979), that such a maximisation problem implies $(\partial e / \partial w)(w/e) = 1$, whilst dropping time subscripts and demand superscripts for convenience. This, known as the Solow condition for efficiency wages, can be rewritten to yield $w = e / (\partial e / \partial w)$. The Solow condition characterises the demand side and as such will be used with the supply information of the previous section to determine both the efficiency pay and the associated level of equilibrium effort that arise. We can now also state the following:

Proposition 3: The firm will always choose an operational wage such that wages affect effort positively.

Proof: See Appendix.

Proposition 3 stands thus in contrast to the discussion following Proposition 2, and the efficiency wage literature on endogenous monitoring levels, which suggested that workers would under certain circumstances want to reduce their effort in response to an increase in wages. Proposition 3 on the other hand states that a firm will always set wages such that the efficiency wage positive correlation between wages and effort holds. There is however no internal conflict between the possibilities that a violation of the condition that underlines Proposition 2 and Proposition 3. For whilst Proposition 2 merely reflects supply responses, both demand and supply factors play a role in Proposition 3, where considerations concerning the conduct of firms when faced with worker behaviour are also included. The firm will in other words chose a wage such that $\partial \tilde{e}_t / \partial \tilde{w}_t > 0$. The operational wage therefore requires

¹² Thus the firm-worker relationship is of a Crawford (1988) type.

$g'(\tilde{e}_t) - \tilde{\eta}_t g(\tilde{e}_t) > 0$, as indicated by expression (A1) in the Appendix, where $\tilde{\eta}_t = (\partial \tilde{p}_t / \partial \tilde{e}_t)(\tilde{e}_t / \tilde{p}_t)$ is the elasticity of detection in time period t with respect to effort in period t . Though not restricted to, this naturally occurs in two cases; when $\tilde{\eta}_t = 0$, as commonly imposed in the literature, and when $\tilde{\eta}_t \leq 1$, the case outlined under Proposition 2. Combining Proposition 2 and Proposition 3, a more complex picture arises than previously acknowledged. Though it is possible that there is range where an increase in wages could reduce a worker's willingness to exert effort, a firm will never chose to operate in such a range, but will instead operate where increases in wages induces more effort from its workforce.

We will in the following return to the literature norm, where $\tilde{\eta}_t = 0$. This is done mainly for simplicity and though we strictly speaking can no longer investigate the endogenous changes across time of the probability of detection should a worker shirk, it avoids overly complex interactions so that we can still determine wage and effort profiles under the two regimes. Restricting the probability elasticity of effort to be zero is furthermore in the spirit of Proposition 3, to which a corollary states that profit maximising efficiency wage firm will operate where the probability elasticity of detection is relatively low.

5.1 Pure Wage Contracts

In this sub-section we investigate the pure wage equilibrium effort in each period by combining both supply and demand behaviour. We can then deduce both the efficiency wage outcome and effort and how these are profiled over time.

Recall from (2) that the wage-effort period-two supply schedule is given by $\tilde{w}_2 = b + g(\tilde{e}_2) / \tilde{p}_2$ in the pure wage case. Total differentiation of this expression implies:

$d\tilde{e}/d\tilde{w}_2 = \tilde{p}_2/[g'(\tilde{e}_2) - (\tilde{w}_2 - b)(\partial\tilde{p}_2/\partial\tilde{e}_2)]$. By combining this with $w = e/(\partial e/\partial w)$ from the Solow (demand side) condition, we derive the profit maximising efficiency wage, \tilde{w}_2^* :

$$\tilde{w}_2^* = \frac{1}{1 + \tilde{\eta}_2^*} \left[\frac{g'(\tilde{e}_2^*)\tilde{e}_2^*}{\tilde{p}_2^*} + \tilde{\eta}_2^* b \right] \quad (14)$$

where $\tilde{\eta}_2^* = (\partial\tilde{p}_2/\partial\tilde{e}_2^*)(\tilde{e}_2^*/\tilde{p}_2^*)$ is the probability elasticity of period-two detection with respect to period-two effort. By substituting expression (14) into $\tilde{w}_2^* = b + [g(\tilde{e}_2^*)/\tilde{p}_2^*]$, we derive an expression for the optimal period-two level of effort:

$$g'(\tilde{e}_2^*)\tilde{e}_2^* - (1 + \tilde{\eta}_2^*)g(\tilde{e}_2^*) = \tilde{p}_2^* b \quad (15)$$

We repeat the above exercise for period-one by totally differentiating the supply expression (4), yielding $d\tilde{e}_1/d\tilde{w}_1 = \tilde{p}_1/\langle g'(\tilde{e}_1) - \{\tilde{w}_1 - b + [(1 - \tilde{p}_2)/\tilde{p}_2]g(\tilde{e}_2)\}(\partial\tilde{p}_1/\partial\tilde{e}_1) \rangle$. Using the Solow condition, we can solve out from this the period-one efficiency wage, w_1^* :

$$\tilde{w}_1^* = \frac{1}{1 + \tilde{\eta}_1^*} \left[g'(\tilde{e}_1^*)\frac{\tilde{e}_1^*}{\tilde{p}_1^*} + \tilde{\eta}_1^* b + \tilde{\eta}_1^* \left(1 - \frac{1}{\tilde{p}_2^*}\right) g(\tilde{e}_2^*) \right] \quad (16)$$

where $\tilde{\eta}_1^* = (\partial\tilde{p}_1/\partial\tilde{e}_1^*)(\tilde{e}_1^*/\tilde{p}_1^*)$ is the elasticity of period-one detection with respect to period-one effort. Replacing this into the left hand side of the supply expression (4) and rearranging then yields the condition for period-one equilibrium effort:

$$g'(\tilde{e}_1^*)\tilde{e}_1^* - (1 + \tilde{\eta}_1^*)g(\tilde{e}_1^*) = \tilde{p}_1^* b + \tilde{p}_1^* \left(1 - \frac{1}{\tilde{p}_2^*}\right) g(\tilde{e}_2^*) \quad (17)$$

A series of results now follows by assuming the elasticity of detection in both periods is, as implicitly is common in the literature, set equal to zero. The probability of detection would therefore not vary across time periods either such that $\tilde{\eta}_1^* = \tilde{\eta}_2^* = 0 \Rightarrow \tilde{p}_1^* = \tilde{p}_2^* = \tilde{p}^*$:

- Proposition 4:*
- (i) *With $\tilde{\eta}_1^* = \tilde{\eta}_2^* = 0$, contractual effort in a pure wage setting is higher in period-two than in period-one;*
 - (ii) *With $\tilde{\eta}_1^* = \tilde{\eta}_2^* = 0$, the wage profile in the pure wage firm is upward sloping.*

Proof: *See Appendix.*

Thus, we have shown that predictions from the agency literature of an upward sloping wage profile is retained. Taking the two parts of the proposition together, it is worth noting that the upward sloping pure wage contract makes the worker more afraid of losing her / his job in period two, which is the period in which the worker is ‘cashing in’ on the firm worker relationship. It therefore makes sense that effort, as well as wages, has a time profile that is upward sloping. Note however, that this result depends heavily on the prevalent literature assumption of constant probability of detection, and were this assumption relaxed there would no longer be certainty about the progression of effort and wages as outlined in the above proposition.

5.2 *The PRP Case*

We now return to the mixed wage and performance related pay case. In this case the firm now realises that it should use total remuneration, rather than just the wage, as its tool to elicit the desired effort from the workers. Thus, we again consider the case where the firm maximises profit, $E\{\pi_t\} = E\{F(e_t N_t) - w_t^* N_t\}$, in each of the two periods.

Using expression (8), the total remuneration-effort period-two (supply) schedule, yields:¹³

$$\frac{\partial \hat{e}_2}{\partial \hat{w}_2} = \frac{\hat{p}_2}{g'(\hat{e}_2) - \lambda(1 - \hat{p}_2) \left[\mathbb{E}\{\theta_{i2} f'(\hat{e}_2)\} - \frac{1}{2} f(0)(\theta_H - \theta_L) \frac{d\hat{p}_2}{d\hat{e}_2} \right] - \frac{\partial \hat{p}_2}{\partial \hat{e}_2} (w_2 - b - \lambda \Delta E_2)} \quad (18)$$

Combining this with the Solow (demand) condition implies the optimal level of ‘efficiency pay’:

$$\hat{w}_2^* = \frac{1}{(1 + \hat{\eta}_2^*)} \left[\frac{\hat{e}_2^*}{\hat{p}_2^*} \Phi_2 + \hat{\eta}_2^* (b + \lambda \Delta E_2) \right] \quad (19)$$

where $\hat{\eta}_2^* = (\partial \hat{p}_2^* / \partial \hat{e}_2^*) \cdot (\hat{e}_2^* / \hat{p}_2^*)$ is the elasticity of period-two detection with respect to period-two effort and where $\Phi_2 = g'(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*) \left\langle \mathbb{E}\{\theta_{i2} f'(\hat{e}_2^*)\} - f(0) [(\theta_H - \theta_L)/2] (d\hat{p}_2^* / d\hat{e}_2^*) \right\rangle$.

By using (8) and (19), we are able to determine the profit maximising equilibrium level of effort:

$$\begin{aligned} & \hat{e}_2^* g'(\hat{e}_2^*) - \hat{p}_2^* b - (1 + \hat{\eta}_2^*) g(\hat{e}_2^*) \\ & = \\ & \lambda(1 - \hat{p}_2^*) \left[\hat{e}_2^* \mathbb{E}\{\theta_{i2} f'(\hat{e}_2^*)\} - \hat{p}_2^* f(0) \left(\frac{\theta_H - \theta_L}{2} \right) \hat{\eta}_2^* - (1 + \hat{\eta}_2^*) \lambda \Delta E_2 \right] - \hat{p}_2^* \hat{\eta}_2^* \lambda \Delta E_2 \end{aligned} \quad (20)$$

We can now draw inferences about the effort profile in the PRP case:

Proposition 5: *Period-two effort is less under a PRP setting than a wage only setting when $\hat{\eta}_2^* = 0$.*

Proof: *See Appendix.*

¹³ We have for notational simplicity denoted $\partial \mathbb{E}\{\theta_{i2} f(\hat{e}_2^*)\} / \partial \hat{e}_2^* \equiv \mathbb{E}\{\theta_{i2} f'(\hat{e}_2^*)\}$. With $dp_i/de_i = \theta'_{ci}(e_i) / (\theta_H - \theta_L)$ and thus $\theta'_{ci}(e_i) = (dp_i/de_i)(\theta_H - \theta_L)$, we therefore have $\partial \mathbb{E}\{\theta_{c2} f(0)\} / \partial \hat{e}_2^* \equiv [\theta'_{c2}(\hat{e}_2^*) / 2] f(0) = (d\hat{p}_2/d\hat{e}_2) [(\theta_H - \theta_L) / 2] f(0)$. Similar notational conventions are followed in relation to period-one.

Since effort and wages are inextricably linked, inferences about second period differences across the two payment regimes can now also be drawn:

Proposition 6: *Period-two remuneration is greater in the wage only setting than in the PRP setting vis. $\tilde{w}_2^* > \hat{w}_2^*$.*

Proof: *Proposition 5 implies that $g(\tilde{e}_1^*) > g(\hat{e}_2^*)$. Proposition 6 then follows by direct comparison of period two wage remuneration under pure wage contracts and PRP from (2) and (8) previously.*

QED.

It appears by viewing propositions 1, 2 and 3 as a whole, that both second period wage and effort are higher in the pure wage case than the PRP case. This may reflect the particular agency issues that arise under the pure wage contract where the wages are necessarily higher later on in the relationship. Thus with wages higher later on in the pure wage arrangement it implies that second period effort is also higher. Such a tendency is weakened when the remuneration is a mix of pure wage payments and payment by performance.

By using (12) we derive:

$$\frac{\partial \hat{e}_1^*}{\partial \hat{w}_1^*} = \frac{\hat{p}_1^*}{\frac{\partial \hat{p}_1^*}{\partial \hat{e}_1^*} \left\{ b - w_1^* + g(\hat{e}_2^*) - \frac{1}{\hat{p}_2^*} [g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*) \Delta E_2] + \lambda \Delta E_1 \right\} + \Phi_1} \quad (21)$$

where $\Phi_1 = g'(\hat{e}_1^*) - \lambda(1 - \hat{p}_1^*) \langle E \{ \theta_{i1} f'(\hat{e}_1^*) \} - f(0) [(\theta_H - \theta_L)/2] (d\hat{p}_1^*/d\hat{e}_1^*) \rangle$. Combining (21)

with the Solow condition yields the first period efficiency remuneration under PRP:

$$\hat{w}_1^* = \frac{1}{1 + \hat{\eta}_1^*} \left\langle \hat{\eta}_1^* \left\{ b + g(\hat{e}_2^*) - \frac{1}{\hat{p}_2^*} [g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*) \Delta E_2] + \lambda \Delta E_1 \right\} + \Phi_1 \frac{\hat{e}_1^*}{\hat{p}_1^*} \right\rangle \quad (22)$$

where $\hat{\eta}_1^* = (\partial \hat{p}_1^* / \partial \hat{e}_1^*) \cdot (\hat{e}_1^* / \hat{p}_1^*)$. Using (22) in conjunction with (12), we derive the period-one equilibrium level of effort:

$$\hat{\eta}_1^* \lambda \Delta E_1 + \Phi_1 \frac{\hat{e}_1^*}{\hat{p}_1^*} = b + g(\hat{e}_2^*) - \frac{g(\hat{e}_2^*)}{\hat{p}_2^*} + \left(\frac{1 - \hat{p}_2^*}{\hat{p}_2^*} \right) \lambda \Delta E_2 + (1 + \hat{\eta}_1^*) \left[\frac{g(\hat{e}_1^*)}{\hat{p}_1^*} - \left(\frac{1 - \hat{p}_1^*}{\hat{p}_1^*} \right) \lambda \Delta E_1 \right] \quad (23)$$

Proposition 7: *Period-two PRP effort under exceeds the first period PRP effort, (i.e. $\hat{e}_2^* > \hat{e}_1^*$) when $\eta_t = 0$.*

Proof: *See Appendix.*

Thus both the PRP and pure wage contracts result under some conditions in upward effort schedules. Note however that we cannot draw conclusions about whether the first period effort under PRP exceeds the first period effort in the pure wage setting, when $\eta_t = 0$ without imposing additional restrictions. Given that we do not know how first period effort levels compare across the two remuneration arrangements, it also follows that first period remuneration are not easily compared either.

Note however that when lifetime effort under both regimes are equalised:

Proposition 8: *Period-one wages are higher under PRP than under the wage only contract, when the net value of the wage only contract does not exceed the PRP contract and the aggregate lifetime effort under the PRP contract is at least as high as the wage only contract.*

Proof: *With the value of the pure wage contract being given as the sum of wages net of the sum of efforts, $\tilde{w}_1^* - g(\tilde{e}_1^*) + \tilde{w}_2^* - g(\tilde{e}_2^*)$, across the two periods and the value of being employed under a PRP arrangement being the sum of total remuneration net of efforts, $\hat{w}_1^* - g(\hat{e}_1^*) + \hat{w}_2^* - g(\hat{e}_2^*)$, over the two periods, we have: $\hat{w}_1^* - g(\hat{e}_1^*) + \hat{w}_2^* - g(\hat{e}_2^*) \geq \tilde{w}_1^* - g(\tilde{e}_1^*) + \tilde{w}_2^* - g(\tilde{e}_2^*)$. With the aggregate lifetime effort condition $g(\hat{e}_1^*) + g(\hat{e}_2^*) \geq g(\tilde{e}_1^*) + g(\tilde{e}_2^*)$ it follows that: $\hat{w}_1^* + \hat{w}_2^* \geq \tilde{w}_1^* + \tilde{w}_2^*$. Since Proposition 6 states that $\tilde{w}_2^* > \hat{w}_2^*$, it follows that $\hat{w}_1^* > \tilde{w}_2^*$.*

QED.

The condition, on which Proposition 8 relies, that the value of the two types of contract are the same, is of interest as it is the condition when the participation constraint binds in both instances. In order for both contracts to be viable and co-exist, the value of these will then have to be equalised. With the additional restriction on aggregate effort of Proposition 8 in place, the wage profile is then flatter under PRP than under the pure wage setting. Under these conditions our main results can be summarised in Figure 3.

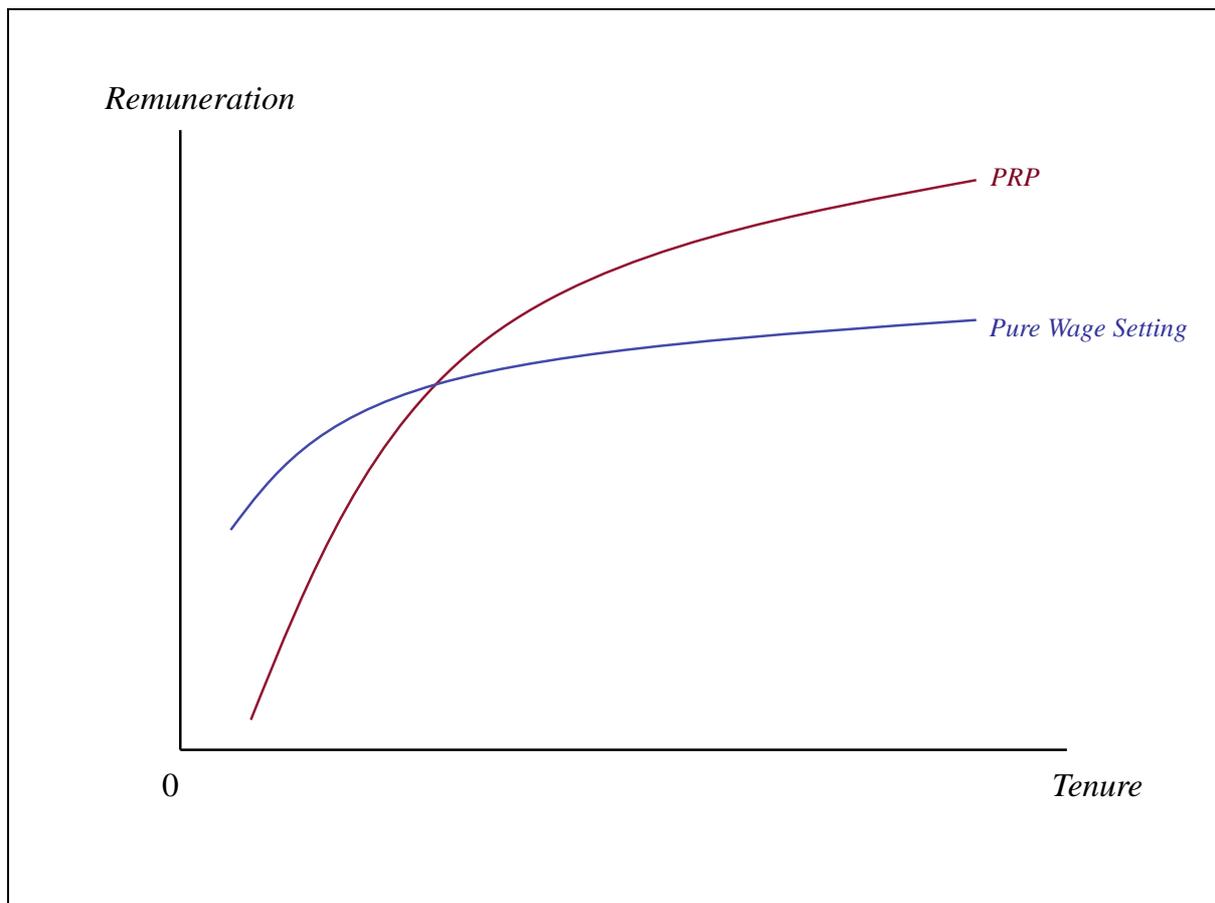


Figure 3: Pure Wage and PRP Remuneration Profiles

The wage profile under PRP is flatter as the numerator under PRP is greater in the beginning and less in the end as compared to the pure wage setting. This is reflected by some straightforward intuition. Period-two remuneration is effectively anchored by the outcome in the spot market. It is easier to satisfy the single period no-shirking condition under PRP than under the pure wage setting as the firm has under PRP an additional instrument or punishment at its disposal. Shirking workers do not only run the risk of being found out and

sacked under PRP, they also face a lower PRP element in their total remuneration package if they do so. Since the firms trade off higher wages against higher effort and it is easier to satisfy the PRP non-shirking condition, the end period remuneration is lower under PRP than under pure wage setting. It in turn influences the second period effort level, which is lower under PRP than under the wage setting regime. Furthermore, there is now also a knock-on effect to the first period where, given the lower remuneration in period two under PRP, it is now relatively more difficult to satisfy the first period *lifetime* non-shirking condition. The PRP firm responds to this difficulty by increasing the compensation and thereby the first period effort level.

The analysis surrounding Figure 3 is nevertheless contingent on the restriction placed on lifetime effort levels. A relaxation of this restriction could potentially, but need not, cause first period remuneration under PRP to be lower than the wage under the wage only contract. It is for instance possible that if effort levels under PRP are sufficiently lower that the remuneration needed to attract workers is such that the PRP compensation schedule is consistently below the ‘pure wage’ compensation schedule. It is then a theoretical possibility that the payment profile is steeper under PRP than under pure wage arrangements.

We must furthermore recognise that Figure 3 is conditioned on the specific assumptions made from Proposition 5 and onwards. Indeed the results in latter part of the paper are premised on setting the elasticities of detection, in both periods, with respect to effort level equal to zero. Matters become more complicated when this constraint is relaxed.. Note also, to derive Proposition 8, that we imposed one further condition, which has the worker’s value of the pure wage contract not exceeding the PRP arrangement value. Such a proviso would of course, in order for both types of payment arrangements to coexist, be required to hold to equality in a market clearing, full-employment economy. Thus the assumption underlying Proposition 8 would under such circumstances be satisfied

automatically. However, with less than full employment, as is the norm in the efficiency wage literature, both types of payments can simultaneously arise, even when the values to the workers across payment regimes are not equalised. Whether or not the value to the worker is greater under the pure wage contract or the PRP scheme is an open question we do not tackle here. Instead, we simply observe that Proposition 8 would hold under the condition stated. However the proposition could be more applicable than it may first appear, for Proposition 8 could easily be augmented to include cases where the clause of having the PRP value never being less than the pure wage contract is relaxed a little. Indeed we could retain the results of Proposition 8 by merely confining the value of pure wage contract to be sufficiently close, that is not ‘too much higher’ than what is derived under the PRP arrangement.

Although it is still early days in terms of empirical work in this area, and although neither study is able to fully control for lifetime effort or contract value, the analyses of Lazear and Moore (1984) and (especially) Brown and Sessions (2006) would appear to lend considerable support to Proposition 8

V. Final Comments

This paper has focused on the relationship between experience-earnings profiles and the degree of worker equity within an enterprise. We extend Lazear and Moore’s (1984) thesis that the nature of the profile is primarily a reflection of agency considerations by focusing not only on those workers with zero or one hundred per cent equity (i.e. salaried and self-employed workers respectively), but also on those with a fractional level of equity *vis.* workers remunerated under some form of *PRP*. Our presumption is that these latter face an intermediate level of agency costs and as such require an intermediate profile. We also extend the efficiency wage literature to endogenise the detection probability of shirkers with respect to effort, and although we recover similar supply side ambiguities with respect to wages and

effort as those found in the endogenous monitoring literature, the supposition made in the more standard efficiency wage literature is demonstrated to be reasonable in the context of our model, as firms never find it optimal to offer wages; where wage increases lead to a reduction in effort.

The nature of the profile has important implications for labour market behaviour. If the slope is primarily a reflection of human capital considerations, then it offers some clue as to the return to on-the-job training and educational investments. If agency considerations are paramount then it raises issues concerning the credibility of long-term employment contracts - firms may have an incentive to fire 'older' expensive, but no more productive, workers. A time-consistency problem may arise, with particular firms unable to recruit younger, less experienced applicants because of their inability to commit not to dismiss them in the future. If the profile, however reflects training, such an incentive-compatibility problem will not arise - older workers will be more productive *ceteris paribus*.

The nature of the earnings profile may also impinge upon quitting behaviour. More experienced 'generally' trained workers will have more manoeuvrability in the labour market than their otherwise similar 'firm-specifically' trained counterparts. But both types may have more options than those older workers whose market rents are primarily a reflection of agency considerations.

Our results might be interpreted as support for the argument that the shape of the experience-earnings profile is primarily a reflection of agency costs rather than the accumulation of human capital through on-the-job training. As such, they highlight important issues pertaining to the credibility of long-term employment contracts, since employers may be tempted to replace experienced workers with less costly, but equally productive, novices. But the latter will not remain 'young' forever, and whether they will be inclined to work for a firm that is unable to guarantee them employment in their dotage is an open question.

Moreover, our findings may help to illuminate a hitherto neglected conduit for the transmission of productivity benefits under collective *PRP* schemes such as profit sharing: if capital markets are imperfect then the same experience-earnings profile would inspire relatively less shirking under a profit-sharing as compared to a salaried contract on account of the lower degree of agency considerations that must be overcome. Alternatively, the same degree of effort may be obtained from risk averse workers *via* a flatter, and therefore less expensive, earnings profile.

References

- Adams, S. J. and J. S. Heywood. (2011). 'Does Deferred Compensation Increase Worker Effort?' *Manchester School*. Forthcoming.
- Akerlof, G. A. (1982) 'Labor Contracts as Partial Gift Exchange.' *Quarterly Journal of Economics*, 97, pp. 543-569.
- Akerlof, G. A. and J. L. Yellen. (1990) 'The Fair Wage-Effort Hypothesis and Unemployment.' *Quarterly Journal of Economics*, 105, pp. 255-283.
- Basu, K. and A. J. Felkey. (2009). 'A Theory of Efficiency Wage with Multiple Unemployment Equilibria: How a Higher Minimum Wage Law Can Curb Unemployment.' *Oxford Economic Papers*, 61(3), pp. 494-516
- Becker, G. S. (1962). 'Investments in Human Capital: A Theoretical Analysis.' *Journal of Political Economy*, 70, pp. 9-49.
- Ben-Porath, Y. (1967). 'The Production of Human Capital and the Life Cycle of Earnings.' *Journal of Political Economy*, 75, pp. 352-65.
- Blinder, A. S. (ed.). (1990). *Paying for Productivity: A Look at the Evidence*. Washington, D.C.: The Brookings Institution.
- Brown, S. and J. G. Sessions. (2006). 'Some Evidence on the Relationship Between Performance Related Pay and the Shape of the Experience-Earnings Profile.' *Southern Economic Journal*, 72(3), pp. 660-676.
- Burdett, K. (1978). 'A Theory of Employee Job Search and Quit Rates.' *American Economic Review*, 68(1), pp. 212-224.
- Carmichael, L. (1990). 'Efficiency Wage Models of Unemployment: One View.' *Economic Inquiry*, 28(2), pp. 269-295
- Crawford, V. P. (1988). 'Long-Term Relationships Governed by Short-Term Contracts.' *American Economic Review*, 78(3), pp. 485-499.
- Danthine, J. and A. Kurmann. (2006). 'Efficiency Wages Revisited: The Internal Reference Perspective.' *Economics Letters*, 90, pp. 278-284
- Frank, R. H. and R. M. Hutchens. (1993). 'Wages, Seniority and the Demand for Rising Consumption Profiles.' *Journal of Economic Behaviour and Organisation*, 21, pp. 251-76.
- Freeman, S. (1977). 'Wage Trends as Performance Displays Productive Potential: A Model and Application to Academic Early Retirement.' *Bell Journal of Economics*, 8, pp. 419-443.
- Fuchs, V. (1981). 'Self-Employment and Labor Market Participation of Older Males.' *Journal of Human Resources*, 18, pp. 339-357
- Goerke, L. (2001) 'On the Relationship Between Wages and Monitoring in Shirking Models.' *Metroeconomica*, 52, pp. 376-390.
- Harris, M. and B. Holmstrom. (1982). 'A Theory of Wage Dynamics.' *Review of Economic Studies*, 49, pp. 315-33.
- Hutchens, R. M. (1987). 'A Test of Lazear's Theory of Delayed Payment Contracts.' *Journal of Labor Economics*, 5, pp. 153-170.
- Hutchens, R. M. (1989). 'Seniority, Wages and Productivity: A Turbulent Decade.' *Journal of Economic Perspectives*, 3(4), pp. 49-64.
- Jacobson, L., and R. J. LaLonde. (1993). 'Earnings Losses of Displaced Workers.' *American Economic Review*, 83(4), pp. 685-709.
- Lazear, E. P. (1979). 'Why is There Mandatory Retirement?' *Journal of Political Economy*, 87 (1979), pp. 1261-84.
- Lazear, E. P. (1981). 'Agency, Earnings Profiles, Productivity and Hours Restrictions.' *American Economic Review*, 71, pp. 606-20.
- Lazear, E. P. (1983). 'Pensions as Severance Pay.' In Bodie, Z. and J. Shoven (Eds.), *Financial Aspects of the U.S. Pension System*. Chicago: University of Chicago Press, p. 57-85.
- Lazear, E. P. (2000). 'The Future of Personnel Economics.' *Economic Journal*, 110, F611-F639.
- Lazear, E. P. and R. L. Moore. (1984). 'Incentives, Productivity and Labor Contracts.' *Quarterly Journal of Economics*, 99, pp. 275-95.
- Loewenstein, G. and N. Sicherman. (1991). 'Do Workers Prefer Increasing Wage Profiles?' *Journal of Labor Economics*, 9, pp. 437-54.
- Manning, A. (2000). 'Movin' On Up: Interpreting the Earnings-Experience Profile.' *Bulletin of Economic Research*, 52, pp. 261-295.
- Medoff, J. L. and K. G. Abraham. (1980). 'Experience, Performance and Earnings' *Quarterly Journal of Economics*, 95, pp. 703-36.

- Medoff, J. L. and K. G. Abraham. (1981). 'Are Those Paid More Really More Productive? The Case of Experience.' *Journal of Human Resources*, 16, pp. 186-216.
- Mincer, J. (1958). 'Investment in Human Capital and Personal Income Distribution.' *Journal of Political Economy*, 66.
- Neumark, D. (1995). 'Are Rising Earnings Profiles a Forced-Saving Investment?' *Economic Journal*, 105, pp. 95-106.
- Polachek, S. and W. Siebert. (1992). *The Economics of Earnings*. Cambridge: Cambridge University Press.
- Ruhm, C. J. (1991). 'Are Workers Permanently Scared by Job Displacements?' *American Economic Review*, 81(1), pp. 319-324.
- Salop, S. (1979). 'A Model of the Natural Rate of Unemployment.' *American Economic Review*, 74, pp. 433-444.
- 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.
- 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*, Vol. 17, pp. 351-376.
- Weiss, Y. (1980). 'Job Queues and Layoffs in Labor Markets with Flexible Wages.' *Journal of Political Economy*, 88, pp. 526-538.
- Weitzman, M. L. and D. L. Kruse. (1990). 'Profit Sharing and Productivity.' In A. Blinder, ed., *Paying for Productivity: A Look at the Evidence*. Washington, D.C.: The Brookings Institution.)
- Wolpin, K. (1977). 'Education and Screening.' *American Economic Review*, 67, pp. 949-958.

Appendix

Proof of Proposition 2:

From (2) it follows that:

$$\frac{\partial \tilde{e}_t}{\partial \tilde{w}_t} = \frac{\tilde{p}_t}{g'(\tilde{e}_t) - \frac{\partial \tilde{p}_t}{\partial \tilde{e}_t} \frac{g(\tilde{e}_t)}{\tilde{p}_t}} = \frac{\tilde{p}_t \tilde{e}_t}{g'(\tilde{e}_t) \tilde{e}_t - \frac{\partial \tilde{p}_t}{\partial \tilde{e}_t} \frac{\tilde{e}_t}{\tilde{p}_t} g(\tilde{e}_t)} = \frac{\tilde{p}_t \tilde{e}_t}{g'(\tilde{e}_t) \tilde{e}_t - \tilde{\eta}_t g(\tilde{e}_t)} \quad (\text{A1})$$

where $\tilde{\eta}_t = (\partial \tilde{p}_t / \partial \tilde{e}_t)(\tilde{e}_t / \tilde{p}_t)$ denotes the elasticity of the probability of detection with respect to effort. Since $g'(\tilde{e}_t) \tilde{e}_t - g(\tilde{e}_t) > 0$ by the convexity of the cost function, the proposition follows.

QED.

Proof of Proposition 3:

With the profit function given by:

$$\pi_t = F(e_t N_t) - w_t N_t \quad (\text{A2})$$

Maximising this with respect to wages yields the following:

$$\begin{aligned} \frac{\partial \pi_t}{\partial w_t} &= \frac{\partial e_t}{\partial w_t} \cdot N_t \cdot F'(e_t N_t) - N_t = 0 \\ \Rightarrow & \\ F'(e_t N_t) &= \frac{1}{\partial e_t / \partial w_t} \end{aligned} \quad (\text{A3})$$

Since marginal product is positive, $\partial e_t / \partial w_t$ is by necessity positive.

QED.

Proof of Proposition 4:

Part (i)

With $\tilde{\eta}_1^* = \tilde{\eta}_2^* = 0$ (i.e. with $\tilde{p}_1^* = \tilde{p}_2^* = \tilde{p}^*$), two relationships follow. From (15) we have:

$$g'(\tilde{e}_2^*) \tilde{e}_2^* - g(\tilde{e}_2^*) = \tilde{p}^* b \quad (\text{A4})$$

From (17) we have:

$$g'(\tilde{e}_1^*) \tilde{e}_1^* - g(\tilde{e}_1^*) = \tilde{p}_1^* b + \tilde{p}_1^* \left(1 - \frac{1}{\tilde{p}_2^*}\right) g(\tilde{e}_2^*) = \tilde{p}^* b - (1 - \tilde{p}^*) g(\tilde{e}_2^*) \quad (\text{A5})$$

Thus

$$g'(\tilde{e}_1^*) \tilde{e}_1^* - g(\tilde{e}_1^*) < \tilde{p}^* b \quad (\text{A6})$$

With the left hand side of both expressions (A4) and (A5) being increasing in the respective time period effort, it follows that $\tilde{e}_1^* < \tilde{e}_2^*$.

QED.

Part ii.

Note that (5) with $\tilde{\eta}_1^* = \tilde{\eta}_2^* = 0$ (and therefore with $\hat{p}_1^* = \tilde{p}_2^* = \tilde{p}^*$) can be written as:

$$\Delta \tilde{w}^* \equiv \tilde{w}_2^* - \tilde{w}_1^* = \left(\frac{1 - \tilde{p}^*}{\tilde{p}^*} \right) g(\tilde{e}_2^*) + \left(\frac{1}{\tilde{p}^*} \right) [g(\tilde{e}_2^*) - g(\tilde{e}_1^*)] \quad (\text{A7})$$

From (A6) we have $g(\tilde{e}_2^*) > g(\tilde{e}_1^*)$. It therefore follows from (A7) that $\Delta \tilde{w}^* > 0$.

QED.

Proof of Proposition 5

The proof follows from (20) which implies that when $\tilde{\eta}_2^* = 0$:

$$g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*)[\hat{e}_2^* E\{\theta_{i2} f'(\hat{e}_2^*)\} - \lambda \Delta E_2] = \hat{p}_2^* b \quad (\text{A8})$$

Note that the concavity of the production function implies:

$$\begin{aligned} & \frac{E\{\theta_{i2} f(\hat{e}_2^*)\} - E\{\theta_{i2} f(0)\}}{\hat{e}_2^*} > E\{\theta_{i2} f'(\hat{e}_2^*)\} \\ \Rightarrow & \Delta E_2 > \hat{e}_2^* E\{\theta_{i2} f(\hat{e}_2^*)\} \\ \Rightarrow & \hat{e}_2^* E\{\theta_{i2} f(\hat{e}_2^*)\} - \Delta E_2 < 0 \end{aligned} \quad (\text{A9})$$

It therefore follows under PRP that:

$$g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*) < \hat{p}_2^* b \quad (\text{A10})$$

When $\tilde{\eta}_2^* = 0$, (15) implies that period-two effort in the pure wage case is given by:

$$g'(\tilde{e}_2^*)\tilde{e}_2^* - g(\tilde{e}_2^*) = \tilde{p}_2^* b \quad (\text{A11})$$

And since the left hand sides of (A10) with (A11) are both increasing in effort, it follows that period-two effort is lower under a PRP contract than it is under a pure wage contract.

QED.

Proof of Proposition 7

The proof is given by contradiction: Assume instead that $\hat{e}_2^* \leq \hat{e}_1^*$. When $\hat{\eta}_1^* = 0$, PRP effort in period-one is determined by (23) which, substituting in for Φ_1 , is given by:

$$\frac{\hat{e}_1^*}{\hat{p}_1^*} \left[g'(\hat{e}_1^*) - \lambda(1 - \hat{p}_1^*) E\{\theta_{i1} f'(\hat{e}_1^*)\} \right] = b + g(\hat{e}_2^*) - \frac{g(\hat{e}_2^*)}{\hat{p}_2^*} + \left(\frac{1 - \hat{p}_2^*}{\hat{p}_2^*} \right) \lambda \Delta E_2 + \frac{g(\hat{e}_1^*)}{\hat{p}_2^*} - \left(\frac{1 - \hat{p}_1^*}{\hat{p}_1^*} \right) \lambda \Delta E_1 \quad (\text{A12})$$

This can be re-expressed as:

$$g'(\hat{e}_1^*)\hat{e}_1^* - g(\hat{e}_1^*) = \hat{p}_1^*b + \hat{p}_1^*g(\hat{e}_2^*) - \left(\frac{1}{\hat{p}_2^*}\right) \left\langle \hat{p}_1^* \left[g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*)\Delta E_2 \right] - \lambda(1 - \hat{p}_1^*) \left[\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* \right] \right\rangle \quad (\text{A13})$$

Second period effort, however, is given by (A8). Subtracting (A8) from (A13) implies:

$$\begin{aligned} & \left[g'(\hat{e}_1^*)\hat{e}_1^* - g(\hat{e}_1^*) \right] - \left[g'(\hat{e}_2^*)\hat{e}_2^* - g(\hat{e}_2^*) \right] + \frac{\hat{p}_1^*}{\hat{p}_2^*} \left[g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*)\Delta E_2 \right] \\ & = \quad \quad \quad (\text{A14}) \\ & \hat{p}_1^*g(\hat{e}_2^*) - \lambda \left\langle (1 - \hat{p}_1^*) \left[\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* \right] + (1 - \hat{p}_2^*) \left[\lambda\Delta E_2 - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^* \right] \right\rangle \end{aligned}$$

Given that that the difference between first two terms on the left hand side of (A14) is non-negative when $\hat{e}_2^* \leq \hat{e}_1^*$, it follows that:

$$\begin{aligned} & \hat{p}_1^*g(\hat{e}_2^*) - \frac{\hat{p}_1^*}{\hat{p}_2^*} \left[g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*)\Delta E_2 \right] \\ & - \quad \quad \quad (\text{A15}) \\ & \lambda \left\langle (1 - \hat{p}_1^*) \left[\Delta E_1 - E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* \right] + (1 - \hat{p}_2^*) \left[\lambda\Delta E_2 - E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^* \right] \right\rangle \geq 0 \end{aligned}$$

Note from (8) that in equilibrium:

$$\hat{w}_2^* - b = \frac{1}{\hat{p}_2^*} \left[g(\hat{e}_2^*) - \lambda(1 - \hat{p}_2^*)\Delta E_2 \right] \quad (\text{A16})$$

Thus we should have:

$$\hat{p}_1^* \left[g(\hat{e}_2^*) - \hat{w}_2^* + b \right] + \lambda \left\langle (1 - \hat{p}_1^*) \left[E\{\theta_{i1}f'(\hat{e}_1^*)\}\hat{e}_1^* - \Delta E_1 \right] + (1 - \hat{p}_2^*) \left[E\{\theta_{i2}f'(\hat{e}_2^*)\}\hat{e}_2^* - \lambda\Delta E_2 \right] \right\rangle \geq 0 \quad (\text{A17})$$

Note, however, that the first term in (A17) is non-positive by virtue of the participation constraint and that the second and third terms are both negative on account of the concavity of production. A contradiction has thus been generated and so it must indeed be the case that $\hat{e}_2^* > \hat{e}_1^*$.

QED.