Newcastle University | Business School

Discussion Paper Series

Economics

4

EFMD EQUIS ACCREDITED

ASSOCIATION OF MBAS ACCREDITED

AACSB ACCREDITED

Athena SWAN Bronze Award

Luck in a Flat Hierarchy: Wages, Bonuses and Noise

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

*Department of Economics
Newcastle University Business School
Newcastle upon Tyne
NE1 4SE
England
Email: john.sessions@newcastle.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: We study bonuses in a flat hierarchy and find effort optimality to be violated within a two worker-type model with noisy performance indicators. Dedicated workers extract informational rents from firms whilst slack worker effort is inefficiently low. Whilst increases in measurement noise reduce the supply-side effort effects of bonuses, they also induce demand-side responses from firms seeking to counteract falling worker exertion by increasing bonuses. Our model helps to explain empirical observations that bonuses are more prevalent and larger in jobs with noisy environments, such as within the financial sector.

Key Words: Bonuses; Banks; Informational Rent.
JEL Classifications: J31, J32, J33, M12.

1. Introduction

How firms reward workers to induce effort has been the subject of labour and personnel economics for many decades. Here it is well known that output contingent payment schemes, such as piece rates and other forms of performance related pay (PRP) in flat organisational structures, and rank order tournament schemes in hierarchical firms, can induce efficient levels of effort.[1] We argue that bonuses can also induce such efficiency and thus provide an explanation for their popularity and widespread use.[2] We illustrate this by modelling a situation where firms can observe neither worker effort nor worker type. In this scenario, appealing remuneration packages must be offered to attract and retain high quality workers. Yet, this leaves firms vulnerable to rent seeking. We demonstrate how the outcome of such opportunistic behaviour is critically dependent on bonus pay.

Our model provides a novel theoretical argument that there is a direct causal link between informational rent, the level of noise within a work environment and the presence and size of bonuses.[3] As such, it sheds light on why workers in jobs with noisy environments and relatively flat hierarchies, like for instance financial traders, receive such a large proportion of their remuneration through bonuses. This is evidenced by Bell and Van Reenen (2014) who find that bonuses are especially widespread in the financial sector. In 2008, whilst 40% of all workers received at least part of their annual earnings through a bonus, in the financial sector this figure rises to 84%. Moreover, bonus payments become increasingly important higher up the income distribution, especially for financial

---

[1] PRP comes in many guises - piece rates, commission, stock options, profit sharing and, importantly, bonuses. However configured, it is widespread within the OECD. Evidence suggests that approximately 20 per cent of U.S. workers and 40 per cent of Dutch and U.K. workers receive some form of PRP - see respectively Gittleman and Pierce (2013), Gielen et al. (2010) and Bell and Van Reenen (2014). The latter find that bonus payments for the lowest nine deciles of wage earners amounts to 2.9% of the total pay of all workers in this group - or a full 7.6% of pay for all workers who receive bonuses. Bryson et al. (2013) similarly conclude that 10-15 per cent of European employees and 40 per cent of Scandinavian and U.S. employees are in receipt of PRP.

[2] We distinguish, by their binary nature, bonuses from PRP more generally. PRP (generally) rewards performance monotonically, whereas bonuses are only triggered if a certain threshold has been achieved.

[3] We concentrate on bonuses within a flat hierarchy in order to crystalise the rent seeking results from our analysis. Thus, whilst we acknowledge Smeets and Warzynski's (2008) empirical finding that bonuses increase with age and job level, we do not investigate the interactions between bonuses and promotions as is done by Ekinci et al. (2018).

sector workers.[4] For all workers outside the top decile, bonuses accounted for 2.9% of total pay. For workers employed in financial services, the comparable figure is 8.6%. Focusing on the top percentile, 83% of all workers received a bonus, with 35% of the total pay for these workers' coming in the form of bonuses. In comparison, financial service workers in the top percentile received 44% of their total pay in bonuses.[5]

Bonuses generally come as either a cash or equity bonus. The latter can be further disaggregated into stock and option holdings and multi-year bonus plans. To qualify for a bonus, the worker must satisfy thresholds as specified by one or more performance measures. Such measures include individual sales and, for CEO's, earnings per share and operating income [Edmans et al. (2017)].[6] Bell and van Reenen (2014) used 2010 pay disclosure data for a sample of London banks to study 1408 top earners in the sector. Their analysis offers specific insight into the extent of bonuses. On average they found that the base salary was outstripped by cash and equity bonuses. Whilst there was some variability across the different banks, the average base wage was £283K with an average associated cash bonus of £285 thousand and an average equity reward of £227 thousand, with deferred cash and equity bonuses of approximately £1.1 million being added to these.

---

[4] Note however that in January 2014, in the aftermath of the financial crisis, the European Union (EU) introduced legislation that capped bankers' bonuses to 100% of the value of fixed pay. The legislation faced a legal challenge from the UK government (which was eventually dropped in November 2014) with the Chancellor of the Exchequer, George Osborne stating: "… these are badly designed rules that are pushing up bankers' pay not reducing it. These rules may be legal but they are entirely self-defeating, so we need to find another way to end rewards for failure in our banks." [*The Guardian*, 20th November 2014)]. Indeed, fixed remuneration has risen since the regulation was introduced - see *Bank of England Quarterly Bulletin*, 4, 2015. Criticism has not been contained to within the UK. France, Ireland, Luxembourg and the Netherlands also resisted the EU drive to implement the bonus cap on smaller financial institutions. [*Financial Times* (29th February 2016)]. Mark Carney, The Governor of the Bank of England, has argued that the cap has: "… the undesirable side effect of limiting the scope for remuneration to be cut back" and has suggested that the cap could be scrapped post-BREXIT [*The Guardian*, 29th November (2017)].

[5] The expansion of bonuses in the US financial sector was, according to Murphy (2013), in large part due to the deregulation of the 1960's. This, he argued, allowed commercial banks to offer an increasing array of services traditionally associated with investment banks: "In order to compete with investment banks in the marketplace, commercial banks also had to compete in the labor market for investment bankers, which meant offering remuneration packages commensurate to those in investment banking. Commercial banks offering investment-banking services faced a growing tension between its traditional commercial bankers – paid high salaries with relatively little performance-based pay – and the professionals in its investment-banking divisions. Ultimately, commercial banks began offering investment-banking-type remuneration for top performers throughout the organization. [Murphy (2013), pp.633-634].

[6] The performance measures usually evaluate absolute performance although some CEO's are also awarded bonuses on the basis of relative performance measures - see Gong et al. (2011).

In what follows we investigate the effects of a simple single-task bonus scheme on worker effort within a noisy environment.[7] We assume that workers are one of two types, differing through their cost of effort, with each receiving a base wage and contingent bonus payments triggered by the attainment of two threshold performance levels.[8] All workers face an idiosyncratic productivity shock such that output depends on both effort and 'luck'. We show that both the level and spread of bonus payments are critically dependent on work environment noise. Increased noise directly reduces worker effort. Firms then respond by widening the gap between the base wage and the bonus in order to re-incentivise its workforce. Thus, these gaps perform a similar effort inducing function to wage spread setting within tournament theory. Indeed, the predictions of our model and the classical Lazear and Rosen (1981) type tournament finding - that increased wage spreads lead to higher effort - are observationally equivalent. They lead, however, to very different conclusions regarding the efficiency of effort exertion. Unlike the use of wage spreads in tournaments, the bonus spreads within our model do not lead to First-Best effort efficiency. [9]

We address in particular the following two questions: First, given their commonality, why does our model generate inefficiencies when efficiency reigns in tournament theory? And second, why does our model and tournament theory produce analogous noise-effort effects? The answers are found in the additional vein of information asymmetry incumbent in our approach. It is the move from perfect information to heterogenous rent seeking workers that generates the inefficiency in our model. This is well known in the literature and effectively precludes the existence of a 'First-Best' structure. Yet, whilst payment structures cannot be efficiency inducing in our model, other results are much more

[7] We concentrate on a single task production process and do not consider multitasking. Matters are considerably more complex when workers must perform several tasks. Holmstrom and Milgrom (1991) suggest that the tendency of workers to relocate effort towards tasks that are relatively more rewarded may induce firms to favour remuneration packages that rely less on PRP. Furthermore, Benabou and Tirole (2016) suggest that competition for talent may lead to aggressive increases in PRP and a subsequent distortion in incentives, resulting in a policy recommendation of bonus caps that may restore efficiency. Such issues are not considered here

[8] Whilst worker type plays a central role in our analysis, we abstract away from the adverse selection concerns found in Greenwald (1986) as well as from the signalling behaviour initially analysed by Spence (1973) and then later by Waldman (1984) and Miklos-Thal and Ullrich (2015).

[9] Lazear and Rosen's (1981) seminal tournament model demonstrates that the application of wage spreads can ensure the optimal exertion of worker effort.

akin to the rank order tournament tradition. These comparable, if not equivalent, results are generated by the informational asymmetries of noise common to both tournament theory and our model.

There is of course always the possibility that firms operating with flat hierarchies might reframe bonuses as promotions, albeit with the same job responsibilities as before. However, as we argue elsewhere, when worker quality is not directly observed firms may prefer privately hidden bonuses to publicly observed promotions in order to limit poaching raids on its workforce by competitor rivals [Sessions and Skåtun (2019)]. Such an argument resonates with the central feature of the market-based tournament literature as formulated by, amongst others, Waldmann (1984), Gibbs (1995), Zabojnik and Bernhardt (2001) and Ghosh and Waldman (2010).[10] Thus, if the external signal of promotions can be avoided by the use of bonuses, then firms might choose the latter as their main incentive mechanism. For some firms, however, promotions are dictated by the distribution of different hierarchy level responsibilities and tasks. Indeed, empirically there is a strong correlation between tasks and promotions. For example, Pergamit and Veum (1999) find that 85% of promoted workers experience an increase in their job responsibilities whilst in Kosteas (2011) 65% of promotions lead to an increase in supervisory tasks. We do not study firms with such hierarchy-specific job responsibilities. We instead focus on flat hierarchy firms where bonuses are typically preferred to promotions to avoid information revelation.

Whilst already an important element of current PRP, recent trends suggest an even wider adoption of bonuses in the future. For whilst bonuses are particularly expedient in flat hierarchies, they are also useful within hierarchical organisations for incentivising workers at the end of the career path or CEO's where no further promotion is possible. Across the OECD, compulsory retirement ages are either increasing or being abolished [OECD (2017)]. Such changes in workplace demographics are likely to have several consequences: First, a more widespread use of bonuses. Second, an increased

[10] The market-based tournament literature also relates to other models of turnover and signalling. For example, 'Up-or-Out' contracts, where workers have an incentive to accumulate general human capital in order to receive outside offers [Waldmann (1990)]. And models where firms endeavour to reduce turnover by making separating wage offers to high- and low-quality workers [Banerjee, and Gaston (2004)].

use of performance management to sanction underperforming workers. And third, relating both to the absolute evaluation associated with bonuses and performance management, an increase in monitoring. Along with other post-Covid changes in work practices, increased monitoring of home working may further facilitate the use of bonuses.[11]

The paper is organised as follows. Section 2 provides some background and context for our study. We use Section 3 to set the scene, simply outlining the full information ('First-Best') and standard, one-dimensional asymmetric information ('Second-Best') models concerning worker type. Yet, in so doing we set the benchmark against which our main results can be measured. Section 4 that then follows is novel and fills a gap that the literature has long ignored - the interaction between bonuses, informational rent and noise in flat hierarchies. This section thus extends our baseline model to consider a 'Third-Best' model of two-dimensional asymmetric information over both worker type and the state of the world; an idiosyncratic shock we refer to as 'luck'. Whilst luck and type are separately unobservable, they may sometimes be collectively deduced. Within this context, we study a contract that offers a base wage with two threshold bonuses. We proceed in this section to investigate the informational rent that high-quality workers can extract and the effect of noise on effort and, ultimately, the bonus spreads within the contract. Final comments are collected in Section 6.

2. Background and Context

The study of worker remuneration under asymmetric information has a long tradition in personnel and labour economics - see Stigler (1962) for an early discussion. We contribute by considering the moral hazard and informational rent issues that arise when workers and firms have incongruent information sets and where the former are remunerated through bonuses.

---

[11] An increased use of bonuses will raise the variability of income across both individuals and groups – for example, age cohorts and income deciles. So, there are implications for income inequality. Whilst interesting and important, we do not address such implications in the present paper.

Our approach, which is to the best of our knowledge unique, is particularly applicable to situations where workers receive a base wage and collect bonuses as they attain pre-arranged performance thresholds. Though there are ample studies that deal empirically with the effect of bonuses on wages and labour market outcomes, relatively few approach the issue from a theoretical perspective.[12]

Whilst much of the theoretical PRP literature deals with piece rate and profit-sharing schemes rather than bonuses, there are some exceptions. MacLeod and Malcomson (1998) compare bonuses with efficiency wages and show that the former prevail when jobs outstrip unemployment. In contrast, efficiency wages are prevalent when unemployment is high. Maestri (2012) concludes that bonus arrangements Pareto dominate efficiency wages, whilst Yang (2008) shows that higher turnover costs result in more bonuses and less efficiency wages.[13] In later work, Maestri (2014) investigates the relationship between termination and bonus decisions, whilst Dur and Tichem (2015) show that altruism shifts incentive mechanisms away from dismissals towards bonuses.

To explain the widespread use of bonuses, Herweg et al. (2010) extend the principal-agent setting of Kőszegi and Rabin (2006) to demonstrate that bonuses are optimal in a Second-Best sense when loss aversion is present. Bakó and Kálecz-Simon (2013), arguing that much of the previous PRP literature has focused on linear payment schedules, study bonus quotas within a differentiated Bertrand duopoly. Similarly, Imhof and Kräkel (2014) discuss the conditions under which firms benefit from offering pooling bonuses to its workforce.

Our model links naturally to tournament theory where relative performance is used to reward the winning (but not the 'losing') workers through promotion. The theory materialises in two main

[12] These empirical studies are many and varied. To mention a few; Green and Heywood (2016) discuss whether bonuses and wages are substitutes. Pouliakas (2010) finds that revoking bonuses has a significant detrimental effect on worker utility but that the positive effect of bonuses on job satisfaction is not permanent. Gibbs and Hendricks (2004) and Cappelli and Conyon (2009) find a positive correlation between appraisal scores and bonuses. Oyer (1998) studies the tendency of workers on end of year bonuses to manipulate their effort levels to suit their own rather than the firm's objective function. The experimental literature has generally found bonus contracts to outperform fixed wage contracts - see for example Fehr et al. (2007). In an experimental paper, Falk et al. (2015) find that dismissal barriers reduce efficiency under fixed wages but that this is completely offset with the introduction of bonuses.

[13] Both MacLeod and Malcomson (1998) and Maestri (2012) rely on subjective evaluations of performance and thus share some communality with our model where effort cannot be observed directly.

forms. In the dominant approach, the 'Classical' tournament model (CTM), the competition prize is determined ex ante and the tournament winner receives a pre-determined wage premium over the loser. See, for example, Lazear and Rosen (1981), Green and Stokey (1983) and Nalebuff and Stiglitz (1983). The alternative form is the 'Market-Based' tournament model (MBTM). Here, promotions are a signal of quality, inducing outside firms to entice the winning workers away from the host firm. In retort, the host firm then ex post offers a higher retention wage and so drives a wedge between the wage of the tournament winners and losers.[14].

In line with both strands of tournament theory, we demonstrate that wider within-firm gross wage dispersions (which in our case include bonuses) are typically associated with higher worker effort. However, our contribution differs from previous work since we do not model promotions but rather a transmission mechanism that relies on informational rent extraction.

In CTM, firms respond to lower worker effort from increased noise by widening the wage spread. In MBTM, in contrast, increased noise reduces the value of a promotion thereby depressing the post wage spread.[15] Whilst we do not want to overstate our case, it should be noted that the effect of noise on bonuses within our model is similar to that of noise on CTM wage spreads. This calls for caution in interpreting such a noise effect as unequivocally identifying CTM.

Our model also links to the literature on heterogenous tournaments where workers of different ability compete against each other. For example, Gürtler and Kräkel (2010) and Gürtler and Gürtler (2015) consider the incentives to exert effort within a MBTM setting. In contrast to CTM, they show here that heterogenous competition leads the weaker player to reduce effort and the stronger player to relax in response.

[14] Ekinci et al. (2018) look at the interaction between bonuses and market-based tournament models and predict that bonuses are used to enhance efficiency whilst being negatively correlated with the level of reward given to a promoted worker

[15] Whether it is more appropriate to model tournaments in a setting where prizes are predetermined, as in CTM, or whether prizes are determined after promotion has taken place, as in MBTM, is an issue that is being increasingly questioned in the literature [Waldman (2013) and DeVaro and Kauhanen (2016)].

Bonuses are likely common when tournaments are impracticable. Whereas the highly incentivised structures that accompany large wage spreads are usually effective in inducing higher effort, they may come with downsides, such as excessive competition between co-workers [Lazear (1989)]. The potential for undue rivalry to hinder teamwork may induce firms to reduce wage spreads in order to dampen the otherwise unfettered competition. Threshold bonuses are attractive to firms since they do not induce such unintended consequences.[16]

Indeed, a key difference between bonuses and tournaments is that the former, but not the latter, are typically dependent on standards and absolute rather than relative performance evaluation. With absolute performance measures, there is no strategic interaction between players. And whilst on the one hand relative performance evaluation has the advantage of being able to filter out common noise, it may have the drawback of undue competition and sabotage. [Lazear (1989)]. The comparative advantages of bonuses and tournaments could thus also be measured in terms of the benefits of reducing industrial politicking versus filtering out common noise.

Bonuses are also an attractive alternative to piece rates within a single- or non-hierarchical structure. When standards are important, the binary nature of bonuses could ensure that thresholds are obtained in contrast to piece rates that monotonically reward low and mediocre levels of output.[17] Furthermore, bonusses do not require continuous monitoring of production, but rather simply for the firm to observe that the threshold standards have been attained. Like tournaments then, bonuses may also be associated with lower costs as compared to piece rates, which require monitoring at all levels of production.

---

[16] Our focus is on the impact of noise on bonus payments rather than why bonuses might be preferred to relative pay. We therefore do not model the choice between relative performance, team performance and individual bonus payments. Kvaløy and Olsen (2012), who do, find that individual PRP rather than team PRP is more common the more are workers able to hold-up the firm. Furthermore, this occurs even when tasks between (indispensable) workers are complementary.

[17] Non-linearities in payment schemes remain to this day an under-researched area in labour economics, a point made some time ago by Prendergast (1999): “Yet many observed contracts are nonlinear, where, for example, discrete bonuses are offered for exceeding some performance threshold. For instance, Kevin J. Murphy (1998) highlights the importance of such bonus contracts for executives. Perhaps the most important form of nonlinearity concerns the threat of being fired, where wages vary little with performance but where poor performance is punished by dismissal. Rather remarkably, the theoretical literature has made little progress in understanding the observed (nonlinear) shape of compensation contracts, despite costs associated with nonlinearities.” [Prendergast (1999), p.15].

A possible undesirable effect of the high-powered and incentive compatible contracts that we consider is an increase in wage inequality [see Lemieux et al. (2009)]. However, we leave fairness issues unexplored as they are not the focus of our model. Neither do we consider the co-existence of low powered / low wage spot market jobs as, for instance, do Esfahani and Mookherjee (1995) who show that firms may provide too few high-powered contracts due to informational rent extraction by workers.

3. Full Information and One-Dimensional Asymmetric Information

As our first benchmark, we outline a model of full information where a firm faces two types of workers. The firm cares solely about profit per worker, $\pi_i = pf(e_i) - w_i$, where p is price and e_i and w_i denote the effort supplied by, and wage paid to, a worker of type i and where $f(e_i)$ denotes the per-worker production function with $df(e_i)/de_i \equiv f'(e_i) > 0$, $d^2 f(e_i)/de_i^2 \equiv f''(e_i) < 0$ and $f(0) = 0$. Type i worker utility is the net surplus between wage and effort cost vis. $u_i(w_i, e_i) = w_i - c(e_i, \delta_i)$, where $c(e_i, \delta_i)$ denotes the type-specific cost of effort function. Workers are either: (i) a 'low-quality' type termed *slack* (s) with a relatively high cost of effort $(\delta_i = \delta_s)$; and (ii) a 'high-quality' type termed *dedicated* (d) with a relatively low cost of effort $(\delta_i = \delta_d < \delta_s)$.[18] The cost function conforms to the standard schedule whereby the marginal cost of effort is positive and strictly convex such that $\partial c(e_i, \delta_i)/\partial e_i > 0$ and $\partial^2 c(e_i, \delta_i)/\partial e_i^2 > 0$. We assume in addition that for a given level of effort the slack have both a higher total and marginal cost of effort vis. $c(e_i, \delta_s) > c(e_i, \delta_d)$ and $\partial c(e_i, \delta_s)/\partial e_i > \partial c(e_i, \delta_d)/\partial e_i$.[19] We then propose:

---

[18] This is similar to the Spence (1973) set-up where some workers have a higher cost of acquiring education. In our model the low-quality worker, who will naturally exert less effort, will do so merely from a cost perspective. No value judgments are intended by using the terms low/ high quality and slack and dedicated, it is simply a reflection of differences in effort.

[19] We assume that higher (than second) order derivatives of the cost function are negligible and therefore set these equal to zero.

Lemma 1. *It is efficient to choose effort (output) that maximises joint welfare. Thus, the optimal First-Best contract is characterised by:*

$$pf'\left(e_s^*\right)=\frac{\partial c\left(e_s^*,\delta_s\right)}{\partial e_s} \tag{1a}$$

$$pf'\left(e_d^*\right)=\frac{\partial c\left(e_d^*,\delta_d\right)}{\partial e_d} \tag{1b}$$

Proof. *Follows from profit maximisation.*

QED.

The expressions in Lemma 1 are included for completeness and provide useful benchmarks conditions that determine the efficient level of production for each type of worker. We also state:

Lemma 2. *The (first-best) equilibrium level of effort for the dedicated is higher than that for the slack such that;* $e_d^* > e_s^*$.

Proof. *Follows from* $\partial c\left(e_i,\delta_s\right)/\partial e_i > \partial c\left(e_i,\delta_d\right)/\partial e_i$ *and Lemma 1*

QED.

We now turn to our second benchmark case of asymmetric information where workers know their own type, but where the firm cannot directly distinguish between types. In this simple one-dimensional asymmetric information case, the wages of the dedicated are always higher than those of the slack to ensure incentive compatibility. Since the firm cannot observe effort/worker type directly, it must instead make the wage a function of observable variable(s), output, itself a function of effort. Thus, the offered contract will specify a high wage to reward high output and a low wage to compensate for the effort needed to produce low output.

To ensure incentive compatibility, the firm must also induce workers to act in accordance to type, that is, to tell the truth:

$$u_d\left(w_d,e_d\right)\equiv w_d-c\left(e_d,\delta_d\right)\geq w_s-c\left(e_s,\delta_d\right)\equiv u_d\left(w_s,e_s\right) \tag{2a}$$

$$u_s\left(w_s,e_s\right)\equiv w_s-c\left(e_s,\delta_s\right)\geq w_d-c\left(e_d,\delta_s\right)\equiv u_s\left(w_d,e_d\right) \tag{2b}$$

Expressions (2a) and (2b) give the incentive compatible remuneration in exchange for effort for the dedicated and the slack respectively.

The firm also needs to ensure workers find their contract sufficiently attractive to participate. This requires the value of the contract to be no worse than the value of leaving the firm for either employment or unemployment elsewhere. In light of asymmetric information, where outside firms have no better information than the workers' current employer, it is taken as given that the workers' outside opportunity is independent of type. Normalising this outside option to zero, we have:

$$u_d\left(w_d,e_d\right) \equiv w_d - c\left(e_d,\delta_d\right) \geq 0 \qquad (3a)$$

$$u_s\left(w_s,e_s\right) \equiv w_s - c\left(e_s,\delta_s\right) \geq 0 \qquad (3b)$$

Expression (3a) and (3b) denote the participation constraints for the respective worker types.[20]

Not all constraints are binding in this asymmetric information case. To see this, let us assume that the dedicated adopt the slack's effort level:

$$\begin{aligned} &u_d\left(w_s,e_s\right) = w_s - c\left(e_s,\delta_d\right) = w_s - c\left(e_s,\delta_s\right) - \Delta c \\ &\Rightarrow \\ &u_d\left(w_s,e_s\right) = u_s\left(w_s,e_s\right) - \Delta c > u_s\left(w_s,e_s\right) \end{aligned} \qquad (4)$$

where $\Delta c = c\left(e_s,\delta_d\right) - c\left(e_s,\delta_s\right) < 0$. Assume then that (3b) is weakly binding such that $u_s\left(w_s,e_s\right) \geq 0$. It now follows from (4) that (3a) is not binding but rather that $u_d\left(w_d,e_d\right) > 0$. This implies the dedicated do better than the slack. In effect, they gather rent. We therefore denote the term $-\Delta c$ in (4) as the informational rent the dedicated worker extracts. Since the participation constraint of the dedicated (3a) is not binding, it now follows that the best the firm can do is to pay the slack workers

[20] We have implicitly assumed $pf\left(e_s\right) \geq w_s$ such that it is in the firm's interest to maintain a relationship with the slack. The firm would want to terminate this relationship if this condition is violated, in which case the firm would prefer the worker did not participate. We do not consider such a degenerate and trivial situation.

as little as possible such that their participation constraint (3b) is binding. Thus, the incentive compatibility constraint for the dedicated (2a) is binding, but that of the slack (2b) is not. The intuitive reason for this is that whereas the firm does not mind the slack mimicking the dedicated by choosing a higher effort level it is much more concerned about the dedicated mimicking the slack and choose too low an effort level.

Even in the instances when a firm cannot observe worker type directly, it is quite reasonable to suppose the firm has some knowledge of the underlying distribution of worker types in the labour market. To reflect this, we apply the simplifying assumption the firm is fully aware of the distribution of workers within the population. We denote the proportion of the workforce who are dedicated as β and the proportion who are slack as $(1-\beta)$. It then follows that the firm's expected profit function may be written as $\mathrm{E}\{\pi\} = \beta\pi_d + (1-\beta)\pi_s$. The contract between the firm and the workers effectively specifies the levels of output that guarantee the high dedicated wage and the lower slack wage. Since there is a one-to-one relationship between effort and output, then the firm is effectively choosing the effort of the dedicated and the slack that maximises this profit subject to the dedicated worker's incentive compatibility constraint and the slack worker's participation constraint:[21]

$$
\begin{aligned}
&\max_{e_d, e_s} \mathrm{E}\{\pi\} = \beta\left[pf(e_d) - w_d\right] + (1-\beta)\left[pf(e_s) - w_s\right] \\
&s.t. \\
&(2a), (3b)
\end{aligned}
\tag{5}
$$

This maximisation programme can be implemented in two simple steps. First, the optimisation problem can be rewritten by utilising the definitions of the utility functions. These can be expressed as $\square w_d \equiv u_d(w_d, e_d) + c(e_d, \delta_d)$ and can be substituted into (5) to yield:

[21] We abstract away from the situation where firms find it optimal to offer a wage (contract) that only attracts the dedicated workers. Though this might be realistic in some instances, the solution is rather trivial. Instead, we focus on the case where it is profitable to offer a menu of wages to attract both sets of workers.

$$\max_{e_d,e_s} \mathrm{E}\{\pi\} = \beta\left[pf(e_d) - c_d(e_d,\delta_d)\right] + (1-\beta)\left[pf(e_s) - c_s(e_s,\delta_s)\right] - \Theta$$

$$s.t. \qquad (6)$$

$$(2a),(3b)$$

where $\Theta = \beta u_d(w_d,e_d) + (1-\beta)u_s(w_s,e_s)$. Second, note the binding incentive compatibility constraint of the dedicated (2a) combined with the binding participation constraint of the slack (3b) implies $u_d(w_d,e_d) \equiv w_d - c(e_d,\delta_d) = -\Delta c$ and $u_s(w_s,e_s) \equiv w_s - c(e_s,\delta_d) = 0$. The optimisation problem (6) thus reduces to:

$$\max_{e_d,e_s} \mathrm{E}\{\pi\} = \beta\left[pf(e_d) - c_d(e_d,\delta_d)\right] + (1-\beta)\left[pf(e_s) - c_s(e_s,\delta_s)\right] + \beta\Delta c \qquad (7)$$

The first-order conditions from the effort choices of the dedicated and slack are:

$$pf'(\hat{e}_d) - \frac{\partial c_d(\hat{e}_d,\delta_d)}{\partial e_d} = 0 \qquad (8a)$$

$$pf'(\hat{e}_s) - \frac{\partial c_s(\hat{e}_s,\delta_s)}{\partial e_s} + \left(\frac{\beta}{1-\beta}\right)\left(\frac{\partial c_s(\hat{e}_s,\delta_d)}{\partial e_s} - \frac{\partial c_s(\hat{e}_s,\delta_s)}{\partial e_s}\right) = 0 \qquad (8b)$$

The first-order conditions (8a) and (8b) imply:

Proposition 1. *(i) The dedicated worker's effort is efficient vis.* $\hat{e}_d = e_d^*$; *(ii) the slack effort is inefficient vis.* $\hat{e}_s < e_s^*$

Proof. *Comparing (8a) with Lemma 1 confirms Proposition 1 part (i). Comparing expressions (8b) and (2b) and using* $\partial c(e_i,\delta_s)/\partial e_i > \partial c(e_i,\delta_d)/\partial e_i$ *confirms Proposition 1 part (ii).*

QED.

Proposition 1 reproduces within our context well known results from the literature – see, for example, Baron and Myerson (1982) and Laffont and Tirole (1986)). In this proposition, the Second-Best outcome is simply a function of the firm's optimising behaviour. Thus, it simply moulds our model to the finding in the literature that, to restrict the rent that the dedicated (as compared to the slack) can extract, the firm shades the slack workers' effort. We now take this framework and go a step beyond previous studies, introducing an additional level of asymmetric information and focusing on the setting of bonuses within firms.

4. Two-Step Bonus Pay and Two-Dimensional Asymmetric Information

It is well known that like tournaments, piece rates also induce efficient effort levels when firms are unable to directly observe effort. [22] In a companion paper we establish the conditions under which bonuses permit a similar outcome [Sessions and Skåtun (2019)]. Specifically, we show that flat hierarchy firms, operating in noisy environments and unable to directly observe worker effort or productivity shocks, can utilise bonuses to achieve first-best effort outcomes. [23]

In the model of Section 3, the only asymmetry of information (in the Second-Best case) was related to worker-type and effort. In this section we retain that asymmetry whilst introducing an additional asymmetry to investigate the Third-Best case. Now, after effort has been chosen, the worker's output is affected by an idiosyncratic shock to productivity that is privately observed by the worker, similar to the effect of noise in the companion paper. We refer to this shock as 'luck' if high or as 'misfortune' if low. Whilst the firm cannot directly observe the idiosyncratic shock or worker effort, it can observe worker output. And worker output can in turn sometimes be used to infer the worker's type or effort level. So, whilst in some cases the firm is unable to ascertain whether the worker's output is due to luck/misfortune or due to effort, worker type is in other cases possible to infer.[24] Indeed, effort / type and luck may in some cases be deduced where: (i) output is so high that

[22] For example, Kim (1997) shows that (under limited liability) a remuneration system combining bonus payments with a wage linear in output (i.e. piece rates) can generate efficient effort outcomes.

[23] The companion paper does not consider informational rent. It instead focusses on homogenous workers receiving a fixed wage, independent of output, and only a single threshold triggered bonus. Whilst resonating with Kim (1997), the paper is more akin to Oyer (2000) who also considers limited liability but who focusses exclusively on threshold bonuses.

[24] By knowing effort, the firm will in equilibrium also be able to identify the worker's type, since the dedicated will exert more effort than the slack.

it can only have arisen from a combination of luck and (high) effort; or (ii) output is so low that the only possibility is that the worker was unlucky and chose low effort. Intermediate levels of output, however, render the firm uncertain as to whether a particular output is the result of luck or effort. Under the asymmetric information circumstances described, the firm can use a simple bonus contract to incentivise its workforce.

In what follows we assume that output is a function of worker effort, *e*, and state dependent 'luck', θ_j, with luck being drawn after effort has been exerted, with no loss of generality, from a uniform distribution with a lower and upper bound θ^L and θ^H respectively such that $\theta_j \in \left[\theta^L, \theta^H\right]$. The uniform distribution greatly simplifies the analysis, and though most of the following propositions carry through, the final Proposition 8 is heavily dependent on the distribution We interpret the distance $\theta^H - \theta^L$ as a measure of noise, where the closer this is to zero the more certain is the environment within which the worker operates. For simplicity, we assume that worker output is given by $y_i = \theta_j f\left(e_i\right)$ with all workers receiving a fixed wage, *w*, and a contingent bonus, *B*, if a given performance threshold is met.[25] This threshold is an arbitrary standard $\bar{y}$, set by the firm, above which workers receive their bonus. It is now apparent that any worker can achieve the standard through exertion and/or luck.

Let us assume that output is a function of worker effort e_i, where $i = d,s$ denotes worker-type, and (uniformly distributed) state dependent luck $\theta_j \in \left[\theta^L, \theta^H\right]$. It follows that, for any given common wage, the effort of a dedicated worker will exceed that of a slack worker. We thus have two output schedules that depend on the draw of luck – see Figure 1 following.

[25] One way of interpreting θ_j is as a type of quasi-price such that $y_i = \theta_j f\left(e\right)$ reflects the overal value or revenue from a worker's output. We could alternatively model worker output with noise as $y_i = \theta_j + f\left(e\right)$ with no significant loss of generality to our results. A key difference between the two specifications is that in our approach luck affects both the level of output and the marginal product of effort.

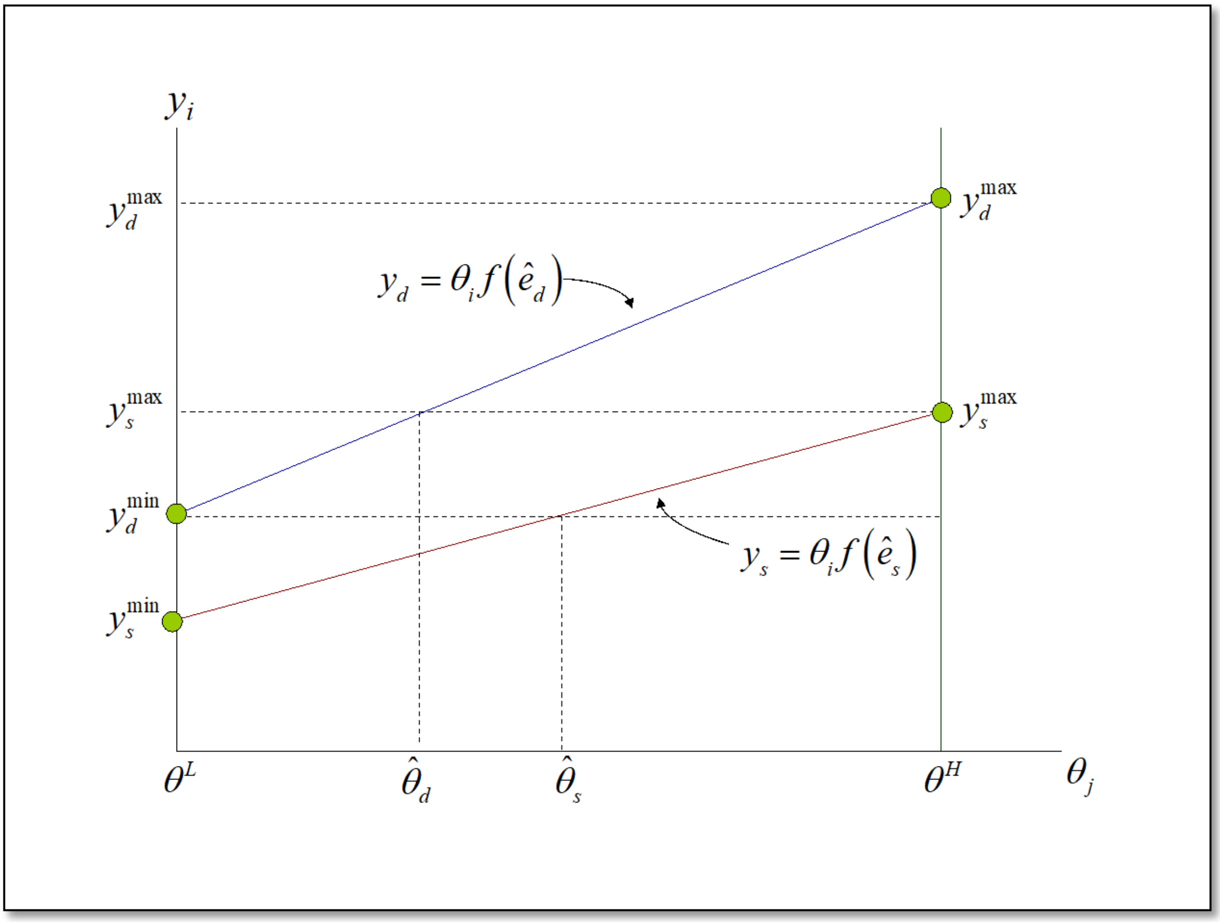

Figure 1: Output Schedules

We now investigate a situation where a worker's overall pay depends on output produced. To the firm, output is informative and can sometimes be used to deduce worker type. In Figure 1 the slack and dedicated equilibrium output levels as a function of luck are given as diverging straight lines - the difference between output levels increases with luck due to our multiplicative specification. With slack effort, e_s, falling short of dedicated effort, e_d, it follows that, for a given realisation of luck, dedicated output exceeds slack output. If output is sufficiently high, then the worker can be revealed to be dedicated. In Figure 1 this occurs whenever output exceeds the highest possible output, $y_s^{\max}$, that a slack worker could produce. Likewise, if output is sufficiently low, then the worker is revealed to be slack. This occurs when output is below the lowest possible output $y_d^{\min}$ that a dedicated worker could produce. In both of these cases the workers will be rewarded according to type. There exists, however, an intermediate band of output within which the firm is unable to distinguish worker-type. We assume that here the firm pays both types equally.

Note in Figure 1 two critical levels of luck or noise: First, $\hat{\theta}_d$, above which the dedicated reveal their type; and second, $\hat{\theta}_s$, below which the slack reveal their type. These critical levels of luck are dependent on both $y_s^{\max}$ and $y_d^{\min}$. We will show in due course that these in turn depend on the equilibrium effort levels of the dedicated and the slack. The critical levels of luck $\hat{\theta}_d$ and $\hat{\theta}_s$ are therefore themselves functions of the equilibrium levels of effort.

We do not consider the fully separating equilibrium in which the slack and the dedicated can always be distinguished. We therefore assume $y_d^{\min} < y_s^{\max}$ in order to avoid the degenerate outcome of complete deduction. Whilst we deal with the more interesting case of partial revelation, it is important to note the restriction this places on the difference in effort between the slack and dedicated. For a given spread of luck, our analysis holds providing this difference is not excessive. This makes sense since it would be easy to spot the dedicated if their effort vastly excceeded that of the slack. It is harder to separate worker types the closer are the effort levels. This more intersting case, where the firm utilises bonus schemes to further incentivise its workforce, is the one we address.

Assuming that all workers accept the contract, then we have the time-sequence set out in Figure 2. As can be seen, nature first reveals type solely to the worker followed by the firm setting the contract. Given this, the worker trhen exerts effort, luck is observed and all outcomes are realised. Note that decisions are only made at Stages 2 and 3 and that we will pin these stages down later on.

In this relatively primitive PRP-bonus model there are three bands of overall pay: (i) a base wage payment to workers revealed to be slack; (ii) a payment comprising the base wage plus a higher bonus payment to workers revealed to be dedicated; and (iii) a payment consisting of the base pay plus a lower bonus to workers whose type the firm is unable to identify – see Figure 3 following. This elementary setting coarsely reflects reality. It is not uncommon to have a lower threshold, below which bonuses are not paid, and a higher cap beyond which bonuses are not increased - see Murphy (1999).[26] In terms of the literature, the range between the two thresholds can be labelled the 'incentive zone'.

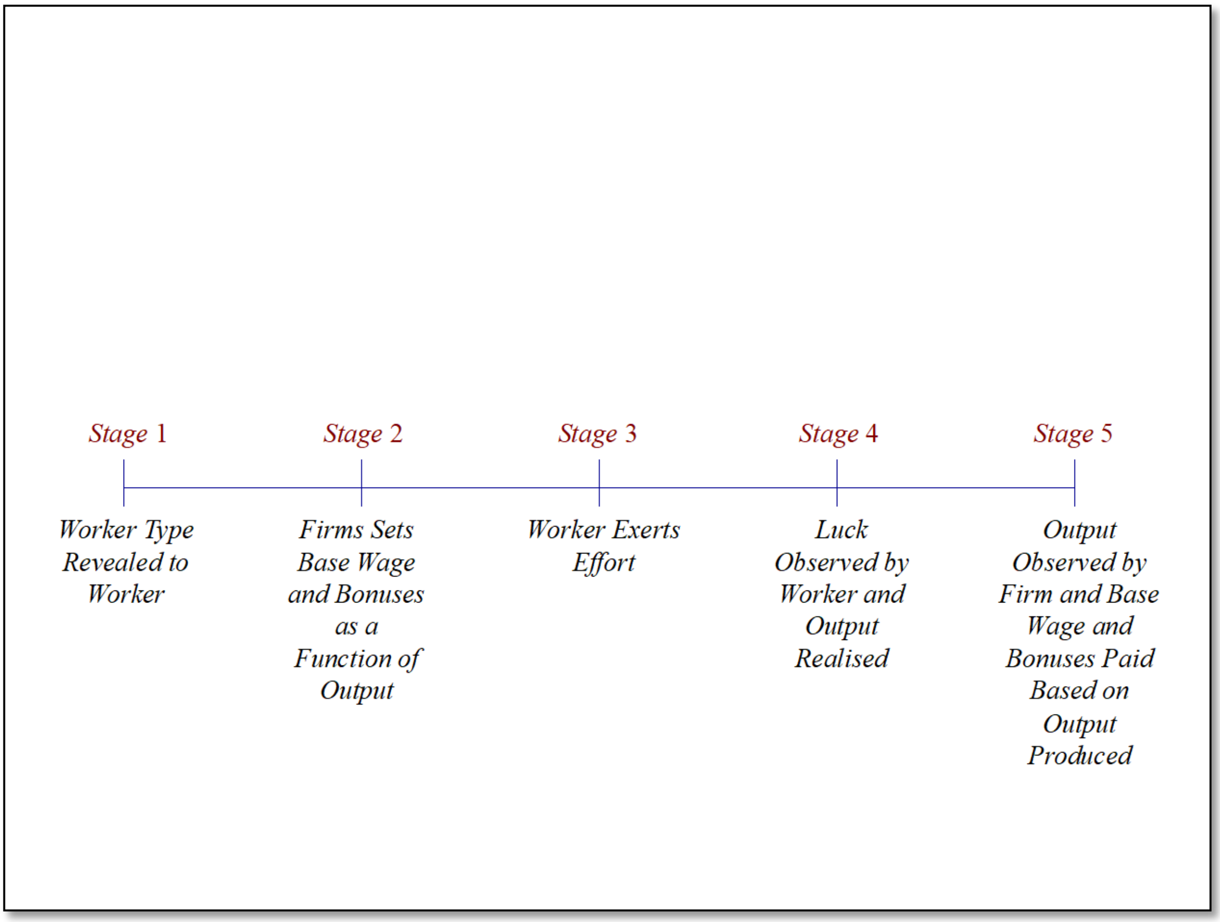

Figure 2: Time Sequence of the Model

The dedicated workers' rent will be shown to increase with the payment to the slack. Thus, this could cause the firm to reduce the payment to the slack workers, and although the dedicated workers choose the optimal levels of effort, the slack workers reduce effort below that which is optimal.[27]

Figure 3 illustrates three regions of output, two of which reveal information about the workers. If $y_i \in \left[y_s^{\min}, y_d^{\min}\right)$, then (recalling Figure 1) the firm can deduce that it is dealing with an unlucky

[26] According to Murphy (1999): "Under the typical plan, no bonus is paid until a threshold performance (usually expressed as a percentage of the performance standard) is achieved, and a 'minimum bonus' (usually expressed as a percentage of the target bonus) is paid at the threshold performance. Target bonuses are paid for achieving the performance standard, and there is typically a 'cap' on bonuses paid (again expressed as a percentage or multiple of the target bonus). The range between the threshold and cap is labelled the 'incentive zone', indicating the range of performance realizations where incremental improvement in performance corresponds to incremental improvement in bonuses."

[27] Note that though probably quite realistic, pay here only alters at pre-determined thresholds where bonuses are triggered.

slack worker who has experienced an idiosyncratic shock below the critical level – i.e. $\theta_s < \hat{\theta}_s = y_d^{\min} / f(\hat{e}_s)$ - and will pay the (slack) worker the base wage w_b.[28]

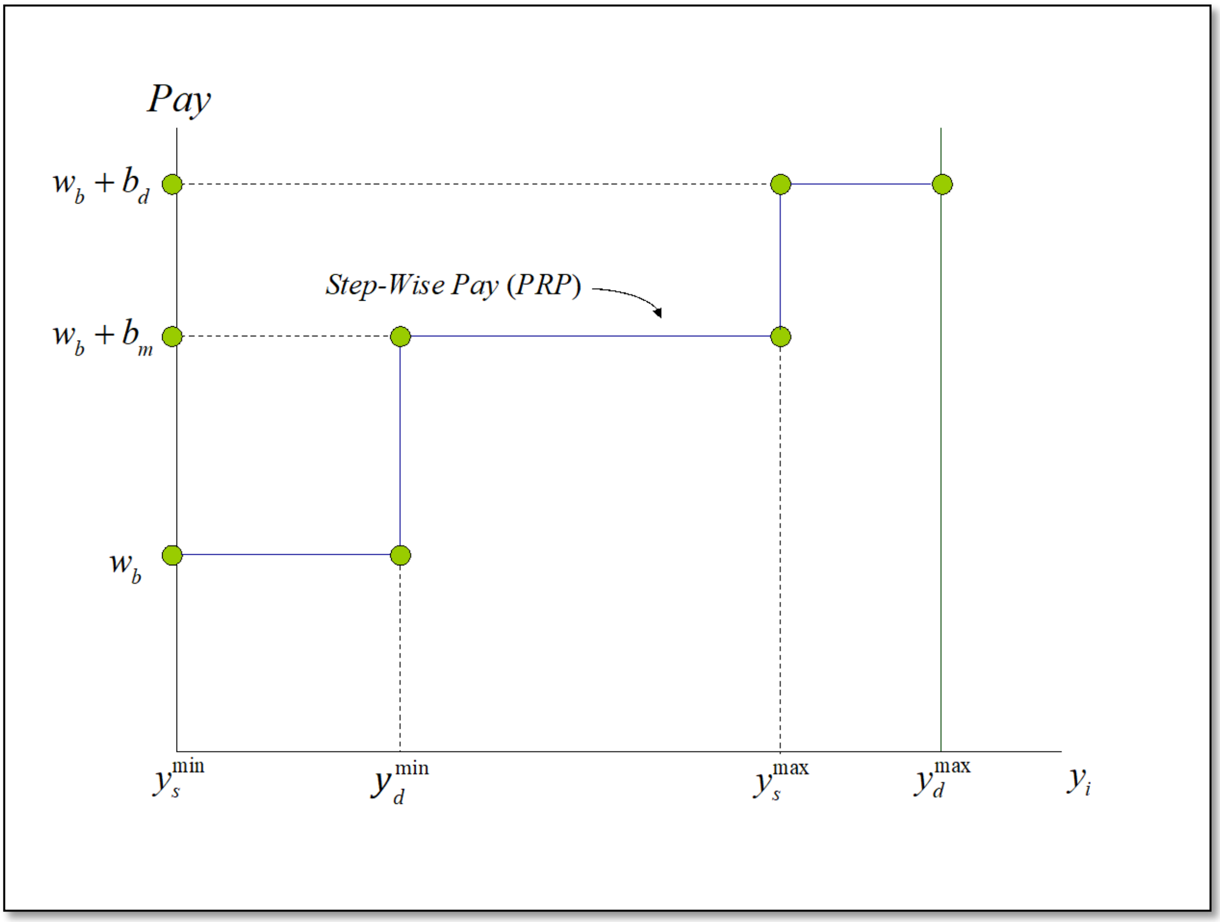

Figure 3: Payment Schedule

At the other extreme is the region $y_i \in \left(y_s^{\max}, y_d^{\max}\right]$. A firm facing a worker who produces output in this region will know that the worker is dedicated and will reward the worker with the gross dedicated wage w_d that comprises the base wage and an additional dedicated bonus b_d vis. $w_d = w_b + b_d$. Outputs associated with the interim region $y_i \in \left[y_d^{\min}, y_s^{\max}\right]$ are more problematic. It is impossible to ascertain whether the worker is dedicated or slack in this region, for the firm cannot determine whether

28 Figure 3 is similar to Figure 11 in Edmans et al. (2017), which describes a 'typical' bonus plan vis: "No bonus is paid until performance reaches a lower threshold, at which point the payoff jumps to the 'hurdle bonus'. On the upside, the bonus is capped at a second threshold. In the 'incentive zone' between the lower and upper threshold, the bonus increases in performance. This increase may be linear … but may also be convex or concave. In the middle of the incentive zone is a 'target' performance level at which a 'target bonus. is awarded." [see Edmans et al. (2017), p. 158].

it is facing an unlucky dedicated worker [i.e. $\theta_d < \hat{\theta}_d = y_s^{\max} / f(\hat{e}_d)$] or a lucky slack worker [i.e. $\theta_s > \hat{\theta}_s = y_d^{\min} / f(\hat{e}_s)$]. We will suppose that the firm remunerates those workers who produce these intermediate levels of production a gross pay $w_m \in (w_b, w_d)$. This comprises a base wage w_b and an additional bonus $b_m \le b_d$ such that $w_m = w_b + b_m$. Unlike Lazear and Rosen (1981), we abstract away from considerations of market structure and conditions when setting bonuses and wages.

Recalling Figure 2, we now analyse the two decision stages of the model vis. *Stage 2 – Firm Sets Base Wage and Bonuses as a Function of Output;* and *Stage 3 – Worker Exerts Effort.* In recursive models of this type, we start with the last decision stage (Stage 3) and use this information to solve for the preceding stage (Stage 2).

Stage 3 - Worker Exerts Effort

Solving the worker's choice of effort first, we consider the payoff to dedicated workers that are given by the difference between the expected wage and the cost of effort. In this case the worker is paid the dedicated wage w_d if $y_i \in (y_s^{\max}, y_d^{\max}]$, which occurs with probability $(\theta^H - \hat{\theta}_d)/(\theta^H - \theta^L)$, and the wage w_m if $y_i \in [y_d^{\min}, y_s^{\max}]$, which occurs with probability $1 - [(\theta^H - \hat{\theta}_d)/(\theta^H - \theta^L)]$. It is evident that the dedicated worker will choose effort so as to maximise the following objective function:

$$\mathrm{E}\{u_d(w_d, w_m, e_d)\} \equiv w_b + \left(\frac{\theta^H - \hat{\theta}_d}{\theta^H - \theta^L}\right) b_d + \left[1 - \left(\frac{\theta^H - \hat{\theta}_d}{\theta^H - \theta^L}\right)\right] b_m - c(e_d, \delta_d) \quad (9a)$$

The relevant objective function for the slack worker is:

$$\mathrm{E}\{u_s(w_s, w_m, e_d)\} \equiv w_b + \left(\frac{\theta^H - \hat{\theta}_s}{\theta^H - \theta^L}\right) b_m - c(e_s, \delta_s) \quad (9b)$$

Since the workers are atomistic and unable to engage in any form of strategic interaction, then the effort of the dedicated and slack affect solely their own critical luck values, $\hat{\theta}_d$ and $\hat{\theta}_s$, respectively. With this in mind, the effort of the dedicated and the slack are derived from:[29]

$$\frac{\partial \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d} = -\frac{\partial \hat{\theta}_d}{\partial e_d}\left(\frac{b_d - b_m}{\theta^H - \theta^L}\right) - \frac{\partial c\left(e_d, \delta_d\right)}{\partial e_d} = 0 \tag{10a}$$

$$\frac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d^2} = -\frac{\partial^2 \hat{\theta}_d}{\partial e_d^2}\left(\frac{b_d - b_m}{\theta^H - \theta^L}\right) - \frac{\partial^2 c\left(e_d, \delta_d\right)}{\partial e_d^2} < 0 \tag{10b}$$

$$\frac{\partial \mathrm{E}\left\{u_d\left(w_s, w_m, e_s\right)\right\}}{\partial e_s} = -\frac{\partial \hat{\theta}_s}{\partial e_s}\left(\frac{b_m}{\theta^H - \theta^L}\right) - \frac{\partial c\left(e_s, \delta_s\right)}{\partial e_s} = 0 \tag{11a}$$

$$\frac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_s^2} = -\frac{\partial^2 \hat{\theta}_d}{\partial e_s^2}\left(\frac{b_m}{\theta^H - \theta^L}\right) - \frac{\partial^2 c\left(e_s, \delta_s\right)}{\partial e_s^2} < 0 \tag{11b}$$

We can now state:

Proposition 2. *Dedicated and slack effort is related to bonusses and bonus spreads as follows:*

$$\frac{de_d}{db_d} = -\left(\theta^H - \theta^L\right)^{-1} \frac{\partial \hat{\theta}_d / \partial e_d}{\dfrac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d^2}} > 0 \tag{12a}$$

$$\frac{de_d}{db_m} = \left(\theta^H - \theta^L\right)^{-1} \frac{\partial \hat{\theta}_d / \partial e_d}{\dfrac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d^2}} < 0 \tag{12b}$$

---

[29] The first- and second-order conditions hold on account of the first- and second-order derivatives for: (i) the dedicated - $\partial \hat{\theta}_d / \partial e_d = -f'(\hat{e}_d) y_s^{\max} / \left[f(\hat{e}_d)\right]^2 < 0$ and $\partial^2 \hat{\theta}_d / \partial e_d^2 = -\left\{f'(\hat{e}_d) f''(\hat{e}_d) - 2\left[f'(\hat{e}_d)\right]^2\right\} y_s^{\max} / \left[f(\hat{e}_d)\right]^3 > 0$; and (ii) the slack $\partial \hat{\theta}_s / \partial e_s = -f'(\hat{e}_s) y_d^{\min} / \left[f(\hat{e}_s)\right]^2 < 0$ and $\partial^2 \hat{\theta}_s / \partial e_s^2 = -\left\{f'(\hat{e}_s) f''(\hat{e}_s) - 2\left[f'(\hat{e}_s)\right]^2\right\} y_d^{\min} / \left[f(\hat{e}_s)\right]^3 > 0$.

$$\frac{de_s}{db_m} = -\left(\theta^H - \theta^L\right)^{-1} \frac{\partial\hat{\theta}_s / \partial e_d}{\frac{\partial^2 \mathrm{E}\left\{u_s\left(w_m, w_s, e_s\right)\right\}}{\partial e_s^2}} > 0 \tag{12c}$$

$$\frac{de_d}{d\left(b_d - b_m\right)} = \left(\theta^H - \theta^L\right)^{-1} \frac{\partial\hat{\theta}_d / \partial e_d}{\frac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d^2}} > 0 \tag{12d}$$

Proof. *Follows directly from expressions (10a)-(10b) and (11a)-(11b)*

QED.

Proposition 2 implies that our model yields a similar prediction to a tournament model. In rank-order tournaments within firms, effort is increasing with the wage spread between winning and losing a contest. In our model a parallel transmission occurs between effort and the spread between receiving bonuses or not. Here, the difference in a dedicated worker's gross wage of receiving the higher or lower bonus must, by necessity, be defined as $w_d - w_m = b_d - b_m$. Likewise, for a slack worker, the difference between the base and gross wage when he receives a bonus is simply $w_m - w_b = b_m$. As is apparent from expressions (12c) and (12d), the increases in the difference in gross wages that arises from either receiving or not receiving the bonuses will increase the effort level of the dedicated and the slack alike.

Expression (12d) shows that the effort of the dedicated worker is higher the greater the gross wage spread between being relatively lucky and being paid the dedicated higher bonus and base wage, and being relatively unlucky and only receiving the lower bonus on top of the base wage. Expression (12c) shows the equivalent condition for the slack worker. Tournament models, which find this result both theoretically [see, for example, Lazear and Rosen (1981)] and empirically [see, for example, Eriksson (1999)] and our model are therefore observationally equivalent when it comes to wage spreads.

The above also indicates that the effect on the wage spreads is different for the dedicated and the slack. With atomistic workers, both $y_s^{\max}$ and $y_d^{\min}$ are invariant to the individual choices of the dedicated and the slack respectively. Since $y_s^{\max} = \hat{\theta}_d f(e_d) \Leftrightarrow \hat{\theta}_d = y_s^{\max} / f(e_d)$ and $y_d^{\min} = \hat{\theta}_s f(e_s) \Leftrightarrow \hat{\theta}_s = y_d^{\min} / f(e_s)$, then $\partial\hat{\theta}_d / \partial e_d = -y_s^{\max} f'(e_d) / f(e_d)^2 < 0$ and $\partial\hat{\theta}_s / \partial e_s = -y_s^{\min} f'(e_s) / f(e_s)^2 < 0$. With $y_s^{\max} > y_d^{\min}$ and $e_d > e_s$, and in light of Footnote 29, it follows that $\left|\partial\hat{\theta}_d / \partial e_d\right| < \left|\partial\hat{\theta}_s / \partial e_s\right|$. Given that we assume zero higher than second-order derivatives, expressions (12c) and (12d) imply that the effect of a given increase in the wage spread is greater for a slack than for a dedicated worker. In other words, the effect of a given increase in the bonus spread is greater the lower the wage. Thus, a second similarity to tournament theory emerges. Consistent with the convexity of the rank / reward structure therein predicted [see, for example, Rosen (1986)], we find that larger wage spreads are required to elicit an increase in effort at higher levels of the wage distribution.

We can also draw inferences regarding the influence of luck, or noise. In tournament models, an increase in noise reduces the return to effort. We measure noise by the spread of the lower and upper idiosyncratic shock variables, $\left(\theta^H - \theta^L\right)$. This noise has two key effects in our model. First, the following proposition applies:

Proposition 3. An increase in noise dampens the effect of bonus spreads on effort.

$$\frac{d^2 e_d}{d\left(b_d - b_m\right) d\left(\theta^H - \theta^L\right)} = -\left(\theta^H - \theta^L\right)^{-2} \frac{\partial\hat{\theta}_d / \partial e_d}{\frac{\partial^2 \mathrm{E}\left\{u_d\left(w_d, w_m, e_d\right)\right\}}{\partial e_d^2}} < 0 \qquad (13a)$$

$$\frac{d^2 e_s}{db_m d\left(\theta^H - \theta^L\right)} = -\left(\theta^H - \theta^L\right)^{-2} \frac{\partial\hat{\theta}_s / \partial e_d}{\frac{\partial^2 \mathrm{E}\left\{u_s\left(w_m, w_s, e_s\right)\right\}}{\partial e_s^2}} < 0 \qquad (13b)$$

Proof. *Follows from a simple use of (10a)-(10b) and (11a)-(11b).*

QED.

The second effect of noise is the direct effect on effort:

Proposition 4. *Increased noise has a direct detrimental impact on effort and reduces the exertion level of both the slack and the dedicated.*

$$\frac{de_d}{d\left(\theta^H-\theta^L\right)}=-\left[\frac{b_d-b_m}{\left(\theta^H-\theta^L\right)^2}\right]\frac{\partial\hat{\theta}_d/\partial e_d}{\frac{\partial^2\mathrm{E}\left\{u_d\left(w_d,w_m,e_d\right)\right\}}{\partial e_d^2}}<0 \tag{14a}$$

$$\frac{de_s}{d\left(\theta^H-\theta^L\right)}=-\left[\frac{b_m}{\left(\theta^H-\theta^L\right)^2}\right]\frac{\partial\hat{\theta}_s/\partial e_d}{\frac{\partial^2\mathrm{E}\left\{u_s\left(w_m,w_s,e_s\right)\right\}}{\partial e_s^2}}<0 \tag{14b}$$

Proof: *Follows from a simple total differentiation of (10a)-(10b) and (11a)-(11b).*

QED.

As can be clearly ascertained from the above, noise has a detrimental impact on effort and tends to reduce the exertion level of both the slack and the dedicated. Increased noise reduces the advantage of the dedicated, whilst it renders the slack relatively more willing to leave matters to luck rather than exert the effort necessary to attain their only feasible bonus, b_m.

Stage 2 – Firm Sets Base Wage and Bonuses as a Function of Output

The incentive compatibility and participation constraints facing the firm at this stage are set out in the Appendix. Given these constraints, we now investigate the informational rent that the dedicated can collect under such circumstances. First, we note that if the dedicated adopt the slack's effort, then:

$$\mathrm{E}\left\{u_d\left(b_m,w_b,e_s\right)\right\}=w_b+\left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)b_m-c\left(e_s,\delta_d\right)$$

$\Rightarrow$

$$\mathrm{E}\left\{u_d\left(b_m,w_b,e_s\right)\right\}=w_b+\left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)b_m-c\left(e_s,\delta_s\right)-\Delta c \tag{15}$$

$\Rightarrow$

$$\mathrm{E}\left\{u_d\left(b_m,w_b,e_s\right)\right\}=\mathrm{E}\left\{u_s\left(b_m,w_b,e_s\right)\right\}-\Delta c>\mathrm{E}\left\{u_s\left(w_m,w_s,e_s\right)\right\}$$

where $\Delta c=c\left(e_s,\delta_d\right)-c\left(e_s,\delta_s\right)<0$ again denotes the informational rent the dedicated worker can collect. The firm's expected profit is thus:

$$\mathrm{E}\left\{\pi\right\}=\beta\left[pf\left(e_d\right)+\Delta c-c\left(e_d,\delta_d\right)\right]+\left(1-\beta\right)\left[pf\left(e_s\right)-c\left(e_s,\delta_s\right)\right] \tag{16}$$

By choosing the bonus levels $\left(b_d,b_m\right)$ the firm maximises this objective function. We derive the first order conditions by noting that whilst b_d only affect the effort of the dedicated, b_m affects the effort of both worker-types. Thus:

$$\frac{\partial\mathrm{E}\left\{\pi\right\}}{\partial b_d}=\beta\Phi_d\left(e_d\right)\frac{de_d}{db_d}=0 \tag{17a}$$

$$\frac{\partial\mathrm{E}\left\{\pi\right\}}{\partial b_m}=\beta\Phi_d\frac{\partial e_d}{\partial b_m}+\left[\beta\left(\frac{\partial c_s\left(\hat{e}_s,\delta_d\right)}{\partial e_s}-\frac{\partial c_s\left(\hat{e}_s,\delta_s\right)}{\partial e_s}\right)+\left(1-\beta\right)\Phi_s\right]\frac{de_s}{db_m}=0 \tag{17b}$$

where $\Phi_i\left(e_i\right)=pf'\left(e_i\right)-\partial c\left(e_i,\delta_i\right)/\partial e_i$, $i=d,s$. Note from (17a) that the effort level of the dedicated is efficient whereas (17b) implies, as before, that the effort level of the slack is inefficient. It is therefore important to emphasise that the informational rent inefficiencies in this rent-seeking case are retained under our particular bonus payment scheme. Thus, whereas we previously argued that some tournament results carry over to our model of informational rent extraction, the efficiency result of CTM theory fails to materialise. So, whilst in our model the firm seeks to use wage spreads to incentivise workers, it cannot avoid inefficiencies. As the dedicated workers' rent depends positively on the pay of the slack, the firm finds it optimal to reduce the slack wage to a level that induces less than First-Best effort.

Stage 3 determines the workers' effort levels and how these levels are affected by both wage profiles and noise, as characterised by expressions (10a) through to (15). Anticipating how such variables affect effort in Stage 3, the firm sets the optimal wage profile in Stage 2, as determined by expressions (17a)-(17b). This enables us now to investigate how profit maximising firms may, by changing their wage structure, counteract the adverse effect of noise on effort. At the extremities of wages, we have:

Proposition 5. *An increase in noise leads to a higher top bonus.*

$$\frac{db_d}{d\left(\theta^H-\theta^L\right)}=\frac{\beta\Phi_d'\left(e_d\right)\cdot\frac{de_d}{d\left(\theta^H-\theta^L\right)}\cdot\frac{de_d}{db_d}}{-\frac{\partial^2 \mathrm{E}\{\pi\}}{\partial b_d^2}}>0 \tag{18}$$

where $\Phi_i'\left(e_i\right)=pf''\left(e_i\right)-\partial^2 c\left(e_i,\delta_i\right)/\partial e_i^2<0.$

Proof. *Follows directly from (17a).*

QED.

Thus, in our model of rent extraction it is clear that noise serves to increase the overall wage spread. To maintain incentives, the firm responds to the lower worker effort that noise ceteris paribus causes by increasing the top bonus.

In terms of the intermediate bonus, we state:

Proposition 6. *An increase in noise leads to a higher intermediate bonus.*

$$\frac{db_m}{d\left(\theta^H-\theta^L\right)}=\frac{\left(1-\beta\right)\Phi_s'\left(e_s\right)\cdot\frac{de_s}{\partial\left(\theta^H-\theta^L\right)}\cdot\frac{de_s}{\partial b_m}}{-\frac{\partial^2 \mathrm{E}\{\pi\}}{\partial w_s^2}}>0 \tag{19}$$

Proof. *Follows directly from (17b).*

QED.

The bonus (or wage spread) for the slack, $b_m = w_m - w_b$, between being detected or not, therefore also increases. What happens to the bonus spread of the dedicated, $b_d - b_m$, remains an open question that we do not directly deal with in this paper. The indeterminacy is rooted in the fact that both $b_d = w_d - w_b$ and $b_m = w_m - w_b$ increase with respect to increases in noise. An increase in the intermediate bonus, b_m, is used by the firm to increase the effort of the slack and to ensure that they participate. This increase has, however, a negative effect on the dedicated workers' effort, which at least to a certain degree can be negated by a higher wage, w_d, at the top. Whilst the effect of noise is indeterminate on the dedicated wage/bonus spread, $w_d - w_m = b_d - b_m$, it is useful to conjecture that this is likely to depend on the underlying distribution of worker types within the firm. Indeed, the effect of noise on the intermediate bonus can be shown to be negatively related to β vis:

Proposition 7. *The effect of noise on the intermediate bonus is negatively related to β.*

$$\frac{d^2 b_m}{d\left(\theta^H - \theta^L\right) d\beta} = \frac{-\Phi_m\left(e_s\right) \cdot \dfrac{de_s}{\partial\left(\theta^H - \theta^L\right)} \cdot \dfrac{de_s}{\partial b_m}}{-\dfrac{\partial^2 \mathrm{E}\{\pi\}}{\partial w_s^2}} < 0 \tag{20}$$

Proof. *Follows directly from (17b).*

QED.

This suggests that this bonus spread may be more likely to increase when there are relatively fewer dedicated versus slack workers employed by the firm. It makes perfect intuitive sense that the firm worries more about the effort levels of the slack when there are many of them. Thus, bonus payments to the slack increase in line with a growth in their numbers within the firm - despite the detrimental effect this may have on the effort level of the dedicated as can be seen from the Proposition 2 effect that a ceteris paribus change in b_m would induce. When instead there are many dedicated workers in the workplace, the firm's attention will be firmly focused on high quality effort, which will ceteris paribus be enhanced by a lowering of the bonus, b_m, as demonstrated by expression (12d).

We have so far not discussed the effect of noise on the base wage. We therefore include a short discussion here. Whilst the effect of noise on bonuses was derived by the use of the first order conditions (10a) and (11a), which are functions of noise and bonuses, it is worth noting that wages do not appear in these first-order conditions. Nevertheless, this does not mean that noise has no effect on base wages. Noise has an indeterminate effect on base wages, working primarily through the participation constraint of the slack worker. Recall that for any given standard and effort level, the critical level of luck is given by $\hat{\theta}_s = y_d^{\min} / f(\hat{e}_s)$. Thus, as increased noise reduces effort, workers need to be luckier in order to attain the bonus. What effect then does an increase in noise have on the utility of the slack? This is not immediately clear. The utility of the slack, as given by (9b) is depends on noise in the following manner:[30]

$$\frac{d\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{d(\theta^H-\theta^L)} = \frac{\partial\left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)}{\partial(\theta^H-\theta^L)}b_m + \frac{\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{\partial b_m}\cdot\frac{\partial b_m}{\partial(\theta^H-\theta^L)} \qquad (21)$$

The latter two right-hand side terms are positive from $\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}/\partial b_m \geq 0$ and Proposition 6. The sign of the first right-hand side term, however, can be positive or negative. We thus consider for illustrative purposes a mean-preserving spread where we note that

30 This follows from the differentiation of (9b) with respect to noise vis:

$$\frac{d\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{d(\theta^H-\theta^L)} \equiv \frac{\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{\partial(\theta^H-\theta^L)} + \frac{\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{\partial b}\cdot\frac{\partial b}{\partial(\theta^H-\theta^L)} + \frac{\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}}{\partial e}\cdot\frac{\partial e}{\partial(\theta^H-\theta^L)}$$

. Expression (21) follows since $\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}/\partial e = 0$ from the envelope theorem and $\partial\mathrm{E}\{u_s(w_s,w_m,e_d)\}/\partial(\theta^H-\theta^L) = \left[\partial\left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)/\partial(\theta^H-\theta^L)\right]b_m$.

$\partial(\theta^H - \hat{\theta}_s)/\partial(\theta^H - \theta^L) = 1/2$. Thus, the mean-preserving change of the first term can be written as:

$$\left.\frac{\partial \mathrm{E}\{u_s(w_s, w_m, e_d)\}}{\partial(\theta^H - \theta^L)}\right|_{Mean\ Preserving} = \frac{\frac{1}{2}(\theta^H - \theta^L) - (\theta^H - \hat{\theta}_s)}{(\theta^H - \theta^L)^2} b_m \tag{22}$$

It is then evident that:

$$\left.\frac{\partial \mathrm{E}\{u_s(w_s, w_m, e_d)\}}{\partial(\theta^H - \theta^L)}\right|_{Mean\ Preserving} \geq 0 \qquad iff \qquad \hat{\theta}_s \geq \tfrac{1}{2}(\theta^H + \theta^L) \tag{23}$$

Let us consider a couple of numerical examples, Case A and Case B, to shed more light on expression (23). For both cases we fix the intial level of noise so that $\theta^H = 30$ and $\theta^L = 10$, impying an expected value of the idiosycratic shock to be given by $\frac{1}{2}(\theta^H + \theta^L) = 20$. The two cases differ in the critical values that the slack worker requires to reach the performance necessary for the bonus to be paid - that is, in the amount of luck the slack worker needs to secure the bonus.To be sure, Case A requires the slack worker to be relatively less lucky than in Case B. We now investigate the impact of a mean-preserving spread as reflected in expressions (22) and (23):

Case A: $\hat{\theta}_s = 15 < 20 = \frac{1}{2}(\theta^H + \theta^L)$

In this case, the probability of receiving the bonus is $(\theta^H - \hat{\theta}_s)/(\theta^H - \theta^L) = 3/4$. Thus, the slack worker has a more than even chance of performing above the bonus threshold. Consider the impact of noise on this probability. Imposing a mean preserving spread such that $\theta^H = 31$ and $\theta^L = 9$ implies:

$$\frac{\theta^H - \hat{\theta}_s}{\theta^H - \theta^L} = \frac{31-15}{31-9} = \frac{16}{22} < \frac{3}{4} \Rightarrow \left.\frac{\partial \mathrm{E}\left\{u_s\left(w_s, w_m, e_d\right)\right\}}{\partial\left(\theta^H - \theta^L\right)}\right|_{Mean\ Preserving} < 0 \tag{24}$$

The probability of obtaining the bonus in this case has fallen for the slack worker. Ceteris paribus, this reduces the utility of the slack. Bonuses are, on the other hand, increasing with noise, to the benefit of the slack, as Proposition 6 demonstrates. Thus, the overall change in the utility of the slack worker from the change in noise, holding the base wage constant, remains uncertain. We cannot, therefore, in this case draw any definite conclusion with respect to how the firm adjusts the base wage in order to retain the binding participation constraint.

We turn now to Case B where, on average, the slack worker does not revceive a bonus:

Case B: $\hat{\theta}_s = 25 > 20 = \frac{1}{2}\left(\theta^H + \theta^L\right)$

In this case the probability of receiving the bonus is $\left(\theta^H - \hat{\theta}_s\right)/\left(\theta^H - \theta^L\right) = 1/4$ such that the slack worker now has a less than even chance of performing above the bonus threshold. Imposing the same mean preserving spread as before such that $\theta^H = 31$ and $\theta^L = 9$ implies:

$$\frac{\theta^H - \hat{\theta}_s}{\theta^H - \theta^L} = \frac{31-25}{31-9} = \frac{6}{22} > \frac{1}{4} \Rightarrow \left.\frac{\partial \mathrm{E}\left\{u_s\left(w_s, w_m, e_d\right)\right\}}{\partial\left(\theta^H - \theta^L\right)}\right|_{Mean\ Preserving} \geq 0 \tag{25}$$

Here, the slack workers who have a less than even chance of attaining the bonus become better off with more noise, since this increases both the bonus and the chance of getting the bonus. This implies that the firm can reduce the base wage and still retain the participation constraint. Intuitively, since the firm does not want to remunerate its workforce more than necessary, then it will under the conditons of Case B reduce the base wage.

Furthermore, going beyond the specific and given the uniform distribution of luck, we can more generally derive:

Proposition 8: *If* $\hat{\theta}_s \geq \frac{1}{2}\left(\theta^H + \theta^L\right)$*, then the base wage is decreasing in noise.*

Proof: *As the participation constraint of the slack holds with equality, that is* $\mathrm{E}\left\{u_s\left(w_s, w_m, e_d\right)\right\} \equiv w_b + \left[\left(\theta^H - \hat{\theta}_s\right) / \left(\theta^H - \theta^L\right)\right] b_m - c\left(e_s, \delta_s\right)$ *=0, then the proposition follows from Proposition 6 combined with expressions (22) and (23).*

QED.

Intuitively, an increase in noise will increase the utility of a slack worker, who has a less than even chance of attaining the bonus, ceteris paribus. Here, as Case B and Proposition 6 demonstrate, a worker with a relatively small probability of being paid a bonus generally welcomes an increase in noise since it increases both the probability and the size of the bonus. The firm can exploit this, since it no longer needs to pay the worker such a high base wage to assure participation. As the probability and size of bonuses increase, then the binding participation constraint is easier to meet. Thus, Proposition 8 provides a sufficient, but not necessary, condition for when it is in the interest of the firm to reduce the wage in response to an increase in noice.

We make one further observation before we conclude. Propositions 5, 6 and 8 together imply that the effect of noise on the wage spread for both worker types is observationally equivalent to CTM but not to MBTM. In the former, increased noise induces a responding firm to set a wider wage spread. In the latter, more noise leads to reduced bidding by competing firms, which in turn compresses the wage spread. The observational equivalence between our model and CTM suggests that caution is needed in using wage spreads per se as the test of which tournament model prevails.

5. Final Comments

In a labour market where rent seeking is pervasive, it stands to reason that evasive actions will be taken. We investigate such behaviour in the presence of a bonus payment regime where the firm uses its pay structure to incentivise and reduce informational rent. In a parallel to tournament theory, we show the firm using bonus spreads to incentivise its workforce. However, unlike the use of wage spreads in tournaments, bonus spreads within our informational rent extraction model do not lead to

First-Best effort efficiency. Nevertheless, a noisier work environment leads to a net reduction in worker effort, inducing the firm to increase the gross wage spread through the use of bonuses.

Our model is especially relevant to the financial sector where bonuses have come under increasing public scrutiny over the last decade, and especially since the financial crisis. The European Union responded in January 2014 by introducing legislation to cap bankers' bonuses to 100% of the value of their fixed pay. But this move has apparently had the inadvertent effect of increasing overall remuneration. Why this might be the case is unclear and reflects the relative ignorance regarding the causes and effects of bonus pay. Indeed, the issue as to why bonuses are so prevalent in the financial sector remains moot.

In this paper we have sought to fill some of the gaps in our understanding. We have set out what we believe is the first plausible model of rent seeking that links the level of noise in job output to the prevalence and extent of bonuses. Yet, though rent seeking and bonuses are connected, it is important to note the bonus payments in our model are not the cause of rent seeking; they are instead its consequence - that is, the tool with which firms seek to control such behaviour. In terms of further research an interesting extension to the work in this paper would be to look more closely at strategic interaction between workers, whether dedicated or slack.

References

Bakó, B. and A. Kálecz-Simon. (2013). 'Quota Bonuses with Heterogeneous Agents.' *Economics Letters*, 119, pp. 316–320.

Banerjee, D. S. and N. Gaston. (2004). 'Labour Market Signalling and Job Turnover Revisited.' *Labour Economics*, 11, pp. 599-622.

Baron, D. P. and R. B. Myerson. (1982). 'Regulating a Monopolist with Unknown Costs.' *Econometrica*, 50, pp. 911-930

Bell, B. and J. Van Reenen. (2014). 'Bankers and their Bonuses.' *The Economic Journal,* 124, pp. F1–F21.

Benabou, R. and J. Tirole. (2016). 'Bonus Culture: Competitive Pay, Screening, and Multitasking.' *Journal of Political Economy*, 124, pp. 305-370.

Bryson, A., R. Freeman, C. Lucifora, M. Pellizzari and V. Perotin. (2013). 'Paying for Performance: Incentive Pay Schemes and Employees' Financial Participation.' In Boeri, T., C. Lucifora and K. Murphy (Eds), *Executive Remuneration and Employee Performance-Related Pay: A Transatlantic Perspective*. Oxford University Press, Oxford.

Cappelli, P. and M. J. Conyon. (2009). 'What Do Performance Appraisals Do?' *Industrial Labour Relations Review*, 20, pp. 1-29.

DeVaro, J. and A. Kauhanen (2016). 'An "Opposing Responses" Test of Classic versus Market-Based Promotion Tournaments.' *Journal of Labor Economics*, 34, pp. 747-779.

Dur, R. and J. Tichem. (2015). 'Altruism and Relational Incentives in the Workplace' *Journal of Economics & Management Strategy*, 24, pp. 485–500.
Edmans, A., X. Gabaix and D. Jenter. (2017). 'Executive Compensation: A Survey of Theory and Evidence.' In Hermalin, B. E. and M. S. Weisbach (Eds), *The Handbook of the Economics of Corporate Governance*, 1, pp. 383-539, Elsevier. London.
Ekinci, E., A. Kauhanen and M. Waldman. (2018). 'Bonuses and Promotion Tournaments: Theory and Evidence.' *The Economic Journal,* 129, pp. 2342-2389.
Eriksson, T. (1999). 'Executive Compensation and Tournament Theory: Empirical Tests on Danish Data.' *Journal of Labor Economics*, 17, pp. 262-280
Esfahani, H. S. and D. Mookherjee. (1995). 'Productivity, Contracting Modes, and Development.' *Journal of Development Economics,* 46, pp. 203-231.
Falk, A., D. Huffman and W. B. Bentley Macleod. (2015). 'Institutions and Contract Enforcement.' *Journal of Labor Economics*, 33, pp. 571-590.
Fehr, E., A. Klein and K. M. Schmidt. (2007). 'Fairness and Contract Design.' *Econometrica*, 75(1), pp. 121–154.
Ghosh, S. and M. Waldman. (2010). 'Standard Promotion Practices Versus Up-or-Out Contracts.' *The Rand Journal of Economics,* 41, pp. 301–325.
Gibbs, M. (1995). 'Incentive Compensation in a Corporate Hierarchy'. *Journal of Accounting and Economics*, 19, pp. 247–277.
________ and W. Hendricks. (2004). 'Do Formal Salary Systems Really Matter?' *Industrial and Labor Relations Review*, 58(1), pp. 71–93.
Gittleman, M. and B. Pierce. (2013). 'How Prevalent is Performance-Related Pay in the United States? Current Incidence and Recent Trends.' *National Institute Economic Review*, 226, pp. 4-15.
Gielen, A. C., M. J. M. Kerhofs and J. C. Van Ours. (2010). 'How Performance Related Pay Affects Productivity and Employment.' *Journal of Population Economics,* 41, pp. 291-301.
Gong, G., L. Y. Li and J. Y. Shin. (2011). 'Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts.' *The Accounting Review*, 86(3), pp. 1007-1043.
Green, C. and J. Heywood. (2016). *'Don't Forget the Gravy: Are Bonuses and Time Rate Complements?' Industrial Relations*, 55, pp. 490-513.
Green, J. R. and N. L. Stokey. (1983). 'A Comparison of Tournaments and Contracts.' *Journal of Political Economy*, 91, pp. 349-364.
Greenwald, B. C. (1986). 'Adverse Selection in the Labour Market.' *Review of Economic Studies,* 53, pp. 325-347.
Gürtler, O. and M. Kräkel. (2010). 'Optimal Tournament Contracts for Heterogeneous Workers.' *Journal of Economic Behavior & Organization*, 75, pp. 180-191.
Gürtler, M. and O. Gürtler. (2015). 'The Optimality of Heterogeneous Tournaments.' *Journal of Labor Economics*, 33(4), pp. 1007-1042.
Herweg, F., D. Müller and P. Weinschenk. (2010). 'Binary Payment Schemes: Moral Hazard and Loss Aversion.' *American Economic Review*, 100, pp. 2451–2477.
Holmstrom, B. and P. Milgrom. (1991). 'Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design.' *Journal of Law, Economics & Organization*, 7, pp. 24-52.
Kim, S. K. (1997). Limited Liability and Bonus Contracts.' *Journal of Economics & Management Strategy,* 6, pp. 899-913.
Kosteas, V. D. (2011). 'Job Satisfaction and Promotions.' *Industrial Relations: A Journal of Economy and Society*, 50, pp 174-194.
Kőszegi, B. and M. Rabin. (2006). 'A Model of Reference-Dependent Preferences.' *The Quarterly Journal of Economics*, 121, pp. 1133-1163.
Kvaløy, O. and T. E. Olsen (2012). 'The Rise of Individual Performance Pay.' *Journal of Economics & Management Strategy*, 21, pp. 301-540.
Laffont, J-J. and J. Tirole. (1986). 'Using Cost Observation to Regulate Firms.' *Journal of Political Economy*, 94, pp. 614-641.
Lazear, E. P. (1989). 'Pay Equality and Industrial Politics.' *Journal of Political Economy*, 9(3), pp. 561-580.
_________ and S. Rosen. (1981). 'Rank-Order Tournaments as Optimum Labor Contracts.' *Journal of Political Economy,* 89, pp. 841-864.
Lemieux, T., B. W. MacLeod and D. Parent. (2009). 'Performance Pay and Wage Inequality.' *The Quarterly Journal of Economics*, 124, pp. 1-49.
Maestri, L. (2012). 'Bonus Payments versus Efficiency Wages in the Repeated Principal-Agent Model with Subjective Evaluations.' *American Economic Journal: Microeconomics*, 4, pp. 34-56
_________ (2014). 'The Efficiency of Bonus-or-Terminate Incentive Schemes under Subjective Evaluations.' *Games and Economic Behaviour*, 87, pp. 412-418.
Miklos-Thal, J. and H. Ullrich. (2015). 'Belief Precision and Effort Incentives in Promotion Contests.' *Economic Journal*, 125(589), pp. 1952-1963.

MacLeod, B. W. and J. Malcomson. (1998). 'Motivation and Markets." *American Economic Review*, 88, pp. 388–411.
Murphy, K. J. (1998). 'Performance Standards in Incentive Contracts.' *Mimeo*, University of Southern California. Subsequently published in: *Journal of Accounting and Economics*, 30(3), December 2000, pp. 245-278.
___________ (1999). 'Executive Compensation.' In Ashenfelter, O. and D. Card (Eds.), *Handbook of Labor Economics*. North-Holland, Amsterdam.
___________ (2013). 'Regulating Banking Bonuses in the European Union: A Case Study in Unintended Consequences.' *European Financial Management*, 19(4), 631-657.
Nalebuff, B. J. and J. E. Stiglitz. (1983). 'Prizes and Incentives: Towards a General Theory of Compensation and Competition.' *Bell Journal of Economics*, 14, pp. 21-43.
OECD. (2017). 'Pensions at a Glance 2017: OECD and G20 Indicators.' *OECD Publishing*, Paris. http://dx.doi.org/10.1787/pension_glance-2017-en
Oyer, P. (1998). 'Fiscal Year Ends and Nonlinear Incentive Contracts: The Effect on Business Seasonality.' *The Quarterly Journal of Economics*, 113, pp. 149–185.
_______ (2000). 'A Theory of Sales Quotas with Limited Liability and Rent Sharing.' *Journal of Labor Economics,* 18, pp. 405-426.
Pergamit, M. R. and J. R. Veum. (1999). 'What is a Promotion?' *Industrial and Labor Relations Review,* 52, pp. 581-601.
Prendergast, C. (1999). 'The Provision of Incentives in Firms.' *Journal of Economic Literature*, 37, pp. 7-63.
Pouliakas, K. (2010). 'Pay Enough, Don't Pay Too Much or Don't Pay at All? The Impact of Bonus Intensity on Job Satisfaction.' *IZA Discussion Paper*, No. 4713
Rosen, S. (1986). 'Prizes and Incentives in Elimination Tournaments.' *American Economic Review*, 76, pp. 701-716.
Sessions, J. G. and J. D. Skåtun. (2019). 'A Bonus Given: Noise, Effort and Efficiency in a Flat Hierarchy.' *Economics Bulletin*, 39, pp. 2527-2532,
Smeets, V. and F. Warzynski. (2008). 'Too Many Theories, Too Few Facts? What the Data Tell Us About the Link Between Span of Control, Compensation and Career Dynamics.' *Labour Economics*, 15, pp. 688-704.
Spence, M. (1973). 'Job Market Signalling.' *Quarterly Journal of Economics*, 87, pp. 355-374.
Stigler, G. J. (1962). 'Information in the Labor Market.' *Journal of Political Economy*, 70(5), pp. 94-105.
Waldman, M. (1984). 'Job Assignments Signalling, and Efficiency.' *The Rand Journal of Economics*, 15, pp. 255-267.
_________ (1990). 'Up-or-out contracts: A Signalling Perspective.' *Journal of Labor Economics*, *8*, pp. 230-250.
_________ (2013). 'Classic Promotion Tournaments Versus Market-Based Tournaments.' *International Journal of Industrial Organization*, 31, pp. 198-210.
Yang, H. (2008). 'Efficiency Wages and Subjective Performance Pay.' *Economic Inquiry*, 46, pp. 179–196.
Zabojnik, J. and D. Bernhardt. (2001). 'Corporate Tournaments, Human Capital Accumulation and the Firm/Size-Wage Relation.' *Review of Economic Studies,* 68, pp. 693-716.

Appendix

The incentive compatibility and participation constraints facing the firm at Stage 2 are given by:

$$\mathbf{E}\left\{u_d\left(b_d,b_m,w_b,e_d\right)\right\} \equiv w_b + \left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)b_d + \left[1-\left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)\right]b_m - c\left(e_d,\delta_d\right)$$

$$\geq \qquad \text{(IC1)}$$

$$\mathbf{E}\left\{u_d\left(b_m,w_b,e_d\right)\right\} \equiv w_b + \left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)b_m - c\left(e_s,\delta_d\right)$$

$$\mathbf{E}\left\{u_s\left(b_m,w_b,e_s\right)\right\} \equiv w_b + \left(\frac{\theta^H-\hat{\theta}_s}{\theta^H-\theta^L}\right)b_m - c\left(e_s,\delta_s\right)$$

$$\geq \qquad \text{(IC2)}$$

$$\mathbf{E}\left\{u_d\left(b_d,b_m,w_b,e_d\right)\right\} \equiv w_b + \left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)b_d + \left[1-\left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)\right]b_m - c\left(e_d,\delta_s\right)$$

$$\mathbf{E}\left\{u_d\left(b_d,b_m,w_b,e_d\right)\right\} \equiv w_b + \left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)b_d + \left[1-\left(\frac{\theta^H-\hat{\theta}_d}{\theta^H-\theta^L}\right)\right]b_m - c\left(e_d,\delta_d\right) \geq 0 \qquad \text{(P1)}$$

$$\mathrm{E}\left\{u_s\left(b_m, w_b, e_s\right)\right\} \equiv w_b + \left(\frac{\theta^H - \hat{\theta}_s}{\theta^H - \theta^L}\right) b_m - c\left(e_s, \delta_s\right) \geq 0 \qquad \text{(P2)}$$