The Empirical Content of Pay-for-Performance

Canice Prendergast*

January 2013

Abstract

Empirical evidence on the effect of pay for performance on output has been scarce. We propose that worker responses to pay for performance plans can be related to their response to a measure of taxes, and suggests an elasticity of output with respect to incentive pay of 0.25 or lower. Furthermore, empirical evidence has also shown little support for the comparative statics of the simplest agency theories. We argue that a reason may be that these tests are not well identified when workers have other reasons for exerting effort.

*University of Chicago Booth School of Business. Email: canice.prendergast@chicagobooth.edu. I am grateful to seminar participants at the NBER, Columbia University, and Duke University for helpful comments. The author declares that he has no relevant material or financial interests that relate to the research described in this paper.

Pay for performance has been a central focus of the economics of organizations for the last three decades. From occupations ranging from schoolteachers to CEOs, it has been both lauded as the solution to misaligned interests and criticized as the source of dysfunctional behavior in the workplace. This enormous literature is premised on two assumptions - first, pay for performance matters for performance, and second, theory has something useful to say about its level and effectiveness. Surprisingly, we know relatively little about whether either is empirically true despite the literature reaching its fourth decade.

Consider the first issue: to what extent does incentive pay affect outcomes? There are studies on individual firms showing how pay affects performance. Such exercises are by necessity narrowly focused as they typically require both personnel data with performance measures, and an experimental setting where incentives exogenously change. This has resulted in relatively few studies so that the fraction of workers for whom we have estimates is negligible. Nor is there much consensus among the existing studies.[1] In sum, little can be said with any certainty about how the average worker would respond to an incentive pay plan.

Most tests of pay for performance use agency theory as its starting point. We argue that the relevant literature may instead be that on labor supply. The first contribution of the paper is to illustrate a relationship, under reasonable assumptions, between the elasticity of output with respect to net marginal income (one minus the tax rate) and the elasticity of output with respect to pay for performance. In our setting, the ratio of the incentive pay elasticity to the tax elasticity is the fraction of all monetary incentives that incentive pay comprises. When incentive pay is the only reason to exert effort, these measures are equal. However, as workers have many potential sources of incentives (promotion, career concerns, etc.), tax responses overstate the response to pay for performance.

The reasons for showing this relationship is that the large literature on responses to taxes can potentially be used to estimate the impact of incentive pay. We use evidence on the relationship between tax rates and taxable income to infer how the average (tax-paying) worker would respond to incentive pay. This suggests an elasticity of output with respect

[1] For example, Lazear, 2000, Bandiera, Barankay, and Rasul, 2005, and Kruse, 1993, find large positive effects of pay on performance. Neal, 2011, surveys the evidence on teachers and argues for significant responses, while the survey of Mahotra et al., 2010, finds little evidence of incentives mattering for doctors, and Mellstrom and Johannson, 2008, and Lacetera et al., 2010, find little evidence of blood donation responding to marginal incentives. Finally, Frey and Jegen's, 2000, survey argue that pay for performance likely reduces output by harming intrinsic motivation while Ariely et al, 2009, find evidence of workers "choking" in high stakes environments such that productivity falls.

to pay for performance of 0.25 or lower, and it is hard to rule out very low responsiveness. These results, which arises in a very standard agency model with two additional assumptions - competitive labor markets, and the absence of income effects - suggest that the impact of incentive pay may be quite limited.

Beyond simply positing a relationship between compensation and performance, one of the primary objectives of agency theory is to inform the frequency and intensity of incentive pay. The second objective of the paper deals with difficulties in empirically identifying such tests. Agency theory addresses settings where workers are rewarded on proxies for their efforts as the efforts themselves cannot be observed. Much of the empirical literature on optimal agency contracts addresses how pay for performance varies as the accuracy of these (unbiased) proxies changes.[2] However, the idea that worse measurement of performance constrains pay for performance has received relatively little supporting evidence (see Prendergast, 2002, for details).

These tests take the standard identification approach of searching for a change in the marginal cost of incentive provision. In the absence of other reasons to exert effort, those exercises are appropriate. However, as with the elasticity exercise above, alternative sources of incentives matter. Specifically, factors that change the marginal cost of pay for performance often simultaneously change the willingness to exert effort for other reasons. If so, identification can fail as the marginal benefit of incentive pay also changes with its marginal cost. We show this for a number of alternative sources of incentives. This renders comparative statics harder to sign and hence existing empirical work difficult to interpret.

1 A Model

An agent is employed by a perfectly competitive firm. The output produced by the agent is given by

$$y = e + a + \epsilon, \tag{1}$$

where e is unobserved effort exerted, a is the agent's ability, and ϵ is an error term. The distribution of ϵ is assumed to be Normal with mean 0 and variance σ^2. There is symmetric uncertainty about the agent's ability, where the distribution of a is assumed by all parties to be Normal with (normalized) mean 0 and variance σ_a^2. The two error terms are uncorrelated.

[2] Another important vein of the agency literature is when proxies are biased (Holmstrom and Milgrom, 1992). While there have been a small number of empirical papers showing how agents respond to biased measures, there is little work addressing how such concerns affect contracts.

The cost of effort for the agent is given by $C(e)$, where $C'(e) > 0, C''(e) > 0$, and $C'(0) = 0$.

Pay-for-Performance The firm potentially offers pay for performance. The agent is paid a wage w equal to

$$w = \beta_0 + \beta y. \tag{2}$$

What matters for the results below is that the agent has some other monetary reason for exerting effort. We focus on career concerns here for illustrative purposes, both as it acts as an alternative incentive, and it has relevant comparative statics below.

Career Concerns The agent cares about how the labor market perceives his ability (Holmstrom, 1999, Gibbons and Murphy, 1992). Specifically, there are future returns based on perceived ability, where that perception is affected by output produced. Output is observed by all. Following the literature, assume that the agent gains an additional monetary reward of $\gamma E(a|y)$, where γ represents the marginal value of perceived ability and $E(a|y)$ is expected ability after y is observed.[3] Note that the same performance measure is used for both sources of incentives.[4]

The agent cares about monetary rewards and effort exerted. If the agent's income is I, his utility is exponential with constant absolute rate of risk aversion r given by

$$U(I, e) = -exp(-r(I - C(e))). \tag{3}$$

Note that this formulation allows us to ignore income effects in the choice of effort.

The agent chooses effort to maximize the sum of monetary rewards minus effort costs. Adding one final issue - a linear tax rate of τ - the objective of the agent is then to maximize

$$E\{-exp\{-r[(1-\tau)[\beta_0 + \beta y + \gamma E(a|y)] - C(e)]\}\}, \tag{4}$$

As is familiar in models of career concerns, expected ability is updated after observing y from its prior of 0 to $E[a|y] = s[y - E(e)]$ where $s = \frac{\sigma_a^2}{\sigma_a^2+\sigma^2}$, and $E(e)$ is the market's expectation of the agent's effort. While the market is not fooled in equilibrium, the agent will exert effort to affect the labor market's (out-of-equilibrium) belief. The objective of the agent is then to choose effort to maximize

$$E\{-exp\{-r[(1-\tau)[\beta_0 + \beta y + \gamma s(y - E(e))] - C(e)]\}\}. \tag{5}$$

[3]This is a simple parameterization of the classic Holmstrom model, where competitive firms compete for workers and are paid their expected productivity. γ would then be determined by both how long the agent garners the value of a good reputation and on how he discounts the future.

[4]As a result, the thought experiment here is to estimate the responsiveness of agents to pay for performance on the same measure on which their other incentives are generated.

2 The Response to Incentive Pay

The relationship between the effort responses to a measure of taxes and to pay for performance is now shown. Consider the agent's incentives, for the moment treating β as exogenous (optimal contracts are considered below). Optimal effort (see the Appendix) is characterized by

$$C'(e^*) = (1-\tau)[\beta + \gamma s]. \tag{6}$$

Intuitively, effort is increasing in pay for performance and the value of a good reputation, and decreasing in tax rates.

Begin by considering how the agent responds to an exogenous increase in pay for performance β. The elasticity of output with respect to incentive pay is given by

$$\zeta_\beta = \frac{\partial y}{\partial \beta}\frac{\beta}{y} = \frac{\beta(1-\tau)}{y(e_i^*)C''(e_i^*)}. \tag{7}$$

Similarly, the elasticity of output with respect to "net marginal income" $1-\tau$ is given by

$$\zeta_{1-\tau} = \frac{\partial y}{\partial(1-\tau)}\frac{(1-\tau)}{y} = \frac{(1-\tau)(\beta+\gamma s)}{y(e_i^*)C''(e_i^*)}. \tag{8}$$

Hence,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \frac{\beta}{\beta+\gamma s} \leq 1. \tag{9}$$

In words, the incentive pay elasticity - measuring local changes in incentive pay - is discounted from the tax elasticity by the fraction of all incentives that incentive pay comprises. (Note that γs represents any monetary reason for exerting effort other than pay for performance.) The two measures are equal only if pay for performance is the sole reason for effort exertion. The reason for higher response to $1-\tau$ is that taxes affects all reasons for exerting effort, while the pay for performance elasticity only captures that one reason for exerting effort. Said another way, the $1-\tau$ elasticity measures the effect of marginally changing all sources of monetary incentives simultaneously. As a result, identifying the effects of a change in taxes on output weakly overestimates the impact of a change in incentive pay.[5]

Our interest is in the effect of a single firm changing its incentive pay. To do this, we are making inferences from an economy-wide change in taxes. The absence of income effects allows us to do this: from (6), the optimal choice of effort is independent of the worker's

[5]Note that the possibility that workers notice pay for performance more than they notice taxes is ruled out. If instead, taxes are ignored by workers (rather like Chetty et al, 2008, for consumption taxes) compared to pay for performance, then the bounds described here may not hold. Also ignored here is the possibility that workers care about whether the government or their employer benefits from their actions.

expected equilibrium utility, so that any market effects caused by tax incidence have no effect on effort choices.

How much do these tax elasticities likely overestimate the response to pay for performance? Ideally one would want empirical estimates of the importance of pay for performance compared to other monetary reasons for exerting effort. Unfortunately, no such estimates exist. Yet from (9), $\frac{\zeta_\beta}{\zeta_{1-\tau}}$ is lowest when β is low, so low pay for performance implies greater overestimates. Yet empirical evidence consistently finds low rates of formal pay for performance in firms.[6] As a result, observed tax elasticities may substantially overstate responses to pay for performance.

2.1 An Application to Taxable Income

The value of this approach is that there is a large literature on the effect of tax changes on supply, which may allow us to make broader statements about the effect of incentive pay than can be done directly. While we could conceivably use this to estimate incentive pay responses for say individual occupations, we now turn to an economy-wide estimate of responsiveness to pay for performance.

Thus far, we have not used the assumption that the market is competitive. Competition allows us to map from the theoretical results on output to observed data on income. Competition only affects β_0, which for any β will be chosen such that expected income for the agent equals expected output, e^*.[7] We therefore begin by considering the responsiveness of taxable income to $1-\tau$ to identify output responses.

There is a large empirical literature estimating the elasticity of taxable income to the tax rate $\zeta_\tau = \frac{-\tau(\beta+\gamma s)}{y(e_i^*)C''(e_i^*)}$. Our interest however is in $\zeta_{1-\tau} = -\frac{1-\tau}{\tau}\zeta_\tau$ so we need to normalize the empirical estimates on ζ_τ by $-\frac{1-\tau}{\tau}$. Do to so, we need relevant marginal tax rates. Most of this literature for the United States is focused on high earners, often those in the top 10%. The relevant normalization used is therefore for this group. Piketty and Saez, 2007, compute a marginal federal tax rate of between 35% and 40% for this group, and with state taxes around 6% on average and also payroll taxes, this implies a renormalization somewhere in the range of -1.25.

[6]For data on the relative infrequency of pay for performance plans, see MacLeod and Parent, 1999, Dube and Freeman, 2010, and Lazear and Shaw, 2009, though Lemieux, McLeod, and Parent, 2009, show that the frequency of pay for performance has been increasing over time. These frequencies typically include firm wide profit sharing plans which offer very low levels of pay for performance.

[7]Remember that expected ability is normalized to 0, so is not included here.

Saez et al, 2009, review the literature on taxable income, and this section owes much to that work. Based on their overview, they conclude the elasticity of taxable income with respect to taxes is in the range of -0.12 to -0.4. If changes in taxable income equalled changes in true income, this would suggest an upper bound for pay for performance of up to 0.5. However, changes in taxable income diverge significantly from changes in income produced, the main component of which appears to be tax avoidance, which we need to eliminate for our purposes. Various strategies have been used to eliminate such tax avoidance:

Elasticities of Taxable Income for Non-Itemizers One possible way of identifying output responses is to consider groups that have less opportunity for tax avoidance. Within the US, the focus has been on those who do not itemize their taxes. First, Saez, 2003, finds an estimate of -0.4 for workers who itemize their taxes, and 0 for those who do not. Similarly, Kopchuk, 2005, only finds positive elasticities for those who have access to tax deductions - those who do not itemize have an estimated elasticity of 0. Saez, 2009, finds similar results. While those who itemize may have inherently higher labor supply elasticities, this suggests substantial influences outside changed inputs to production. Saez also notes that the lion's share of the tax elasticity arises from the self-employed, whose opportunities for tax avoidance are much greater than for other workers.[8] Estimates for the non self-employed are in the range of -0.08.

In summary, those who have limited opportunities for tax avoidance appear to have taxable income elasticities below -0.1.

Wider Measures of Income Another possible avenue is to consider measures of income that are less taxed and to see if elasticities fall. Broad income includes non-taxed items such as IRA contributions. The elasticity of broad income with respect to tax rates is lower than taxable income, with estimates in the range of -0.12 (Gruber and Saez, 2002). Giertz, 2007, finds a slightly larger number of -0.15.

Different Time Periods The ability to avoid taxes has increased in the US over time. Given this, it is notable that elasticities were much lower in the Kennedy era, close to the low end of these estimates (Saez et al, 2009).

Different Countries Opportunities for avoiding taxes are considerably lower in some other developed countries. Relevant elasticities are also lower. In Canada, where deduction

[8]Of course, they also have more flexibility over hours decisions too.

opportunities are limited, Silamaa and Veall, 2001, finds an estimate of -0.25. Estimates for Sweden are in the -0.2 to -0.4 range (Selen, 2002, Hansen, 2007), while Norway's estimates are close to 0 (Aarbu and Thoreson, 2001). Finally, Kleven and Schultz, 2009, find modest estimates for Denmark.[9]

On the basis of this, -0.2 seems if anything a generous estimate of the elasticity of the tax rate on income produced. Normalizing by $-\frac{1-\tau}{\tau} = -1.25$ and imposing the assumption of competitive markets yields an upper estimate of roughly 0.25 for the elasticity of $1-\tau$ on output. If formal pay for performance was the only monetary reason for exerting effort, this also would be the upper estimate of the elasticity of output with respect to incentive pay.

Yet workers clearly have many other reasons for exerting effort, among them career concerns, a desire to be promoted, and direct observation of effort (input monitoring). As a result, 0.25 may be a substantial overestimate, especially given the low levels of observed pay for performance described above.[10] As a trivial example, if formal pay for performance constitutes a third of all marginal incentives for the average worker considered here (which seems high) and the true elasticity of output with respect to $1-\tau$ is 0.25 (which again seems high), then the incentive pay elasticity is only 0.08. As a result, the impact of incentive pay on performance may be very limited.

To summarize, this section makes a simple observation - as taxes change the marginal returns to worker inputs, they can potentially inform how those same workers would respond to pay for performance. Both carry out a similar conceptual exercise: using exogenous variation to plot the marginal cost curve for inputs to the production process. As the data requirements for precise studies on pay for performance are difficult to attain, which would explain the paucity of work on the issue, this may be an alternative avenue to understanding the role of incentive pay. At a minimum, it offers a challenge to those who believe that pay for performance effects are large, namely, if workers don't respond much to taxes, why should they respond significantly to pay for performance? Both affect the relationship between effort and pay in the similar way.

[9]With higher marginal tax rates in these countries than in the US, the normalization factor would be lower here.

[10]The tax literature also considers the possibility that firms change the composition of pay, where tax increases lead to a shift into non-taxed benefits such as medical care and pensions. This is conceptually identical to the tax shifting argument above and if this effect is important, would reduce the pay for performance elasticity further.

2.2 Potential Sources of Empirical Bias

The results above rely on the assumptions of competitive markets and an absence of income effects. These are used to impute the effect of an individual firm changing incentive pay from market wide changes in taxes, by allowing us to ignore certain "market effects". Such market effects arise in two ways.

First, competitive markets imply that all tax incidence is on workers. As a result, any change in average income above is due to changed inputs. As data on income is more easily attained than data on output, this has an obvious appeal. This mapping is not true when incidence is shared, as changes in income caused by tax changes deviate from changes in output. The theory above still holds, but care must be taken to make productivity inferences from income changes.[11]

There are a number of responses to argue that this is unlikely to be important. First, it takes time for costs imposed on workers to become incident on firms, and (as in the labor supply literature) we address this by only considering short run estimates, measuring effects before such incidence on firms presumably arises. Second, the small empirical evidence on incidence finds support for incidence on workers even over the long run. For example, Gruber, 1997, shows complete incidence of payroll tax changes in Chile on workers, so incidence issues even over the long run may not be an issue.

A final method of addressing whether incidence matters is not to consider income, but rather directly identify how labor inputs change with taxes. To this end, consider how taxes affect the only observable component of inputs, namely hours worked. The response of hours to taxes remains a contentious issue. Different methodologies have been used to provide both macro and micro estimates on both extensive and intensive margins. Our interest is primarily on the intensive margin of a given worker changing effort on the job.[12] On the intensive margin, estimated responses are modest. While some estimates of elasticities reach -0.3 (Chetty, 2012, and Keane, 2011), most now argue that responses are in the -0.1 range. See Saez, 2012, for details. As such, these estimates on hours are consistent with the low responses in taxable income above.

The second "market effect" derives from income effects. Consider the effect of a tax increase. With incidence on the worker, his income (holding effort constant) falls. As pay for

[11]It is not necessary that the agent receive all of output. What matters is that the ratio of taxable income to output is independent of the tax rate. Problems arise when this is not the case. For example, consider the case where the elasticity of worker income with respect to $1-\tau$, holding output constant, is $z<0$. Then elasticity of income with respect to $1-\tau$ understates the elasticity of output with respect to $1-\tau$ by $-z$.

[12]Results on the extensive margin are discussed below.

performance has no such income effect, and the income effect moves in the opposite direction to the substitution effect, this could violate the upper bound exercise above. Income effects could therefore be relevant with a different formulation of the utility function. However, this is unlikely to be a substantive issue. Many papers have estimated income effects in the United States, and estimates tend to be very low (see Blundell and MaCurdy, 1999, and Gruber and Saez et al, 2002, for details), so empirically this is not a large concern.

3 Agency Concerns

Agency theory addresses problems that arise when efforts are not observable. Here we deal with the two primary themes of the literature. First, the agent's performance above is measured with (unbiased) error. Such errors typically constrain optimal incentive pay. Second, we assume above that incentive pay is based on true output. In many settings, this is unlikely to be true, as many aspects of performance cannot be included in an incentive contract. We deal with each in turn.

3.1 Optimal Contracts

From (9), the ratio of the incentive pay elasticity to the tax elasticity is the fraction of all incentives that pay for performance comprises. As a result, agency concerns affect inferences from tax elasticities to the extent that they change the *relative* importance of incentive pay.

First consider the effect of unbiased measurement error. Noisy performance measures affects the choice of the preferred β, so consider optimal contracts. This is an application of Gibbons and Murphy, 1992, and in the Appendix is shown to yield optimal incentive pay of

$$\beta^* = \frac{1}{1 + r(\sigma^2 + \sigma_a^2)C''} - \gamma s. \tag{10}$$

The optimal contract is discounted below full residual claimant ($\beta^* = 1$) by the risk aversion of the agent and his other incentives. Furthermore, when contracts are optimally chosen,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = 1 - \gamma s(1 + r(\sigma^2 + \sigma_a^2)C''). \tag{11}$$

This allows us to characterize the overestimate of incentive pay elasticities from tax data in terms of the parameters that generate agency distortions, namely, risk aversion and measurement error. Note first that it is not necessary that agency be costly for $\frac{\zeta_\beta}{\zeta_{1-\tau}} < 1$: when the worker is risk neutral, unobservability of e is costless yet $\frac{\zeta_\beta}{\zeta_{1-\tau}} = 1 - \gamma s \leq 1$. As a result, these results hold absent agency costs.

Two parameters typically affect optimal contracts: risk aversion and measurement error. Above we noted that $\frac{\zeta_\beta}{\zeta_{1-\tau}}$ is decreasing in β. It is tempting to argue from this that greater agency concerns - higher risk aversion or worse measurement - reduces $\frac{\zeta_\beta}{\zeta_{1-\tau}}$. This is not necessarily true, as what matters is how they affect pay for performance relative to other incentives. (This is the reason for including career concerns above, as it has relevant comparative statics that affect the results.) To see this, consider two important parameters in (11):

Risk Aversion The right hand side of (11) is decreasing in r. As a result, agency concerns generated through risk aversion result in the tax elasticity more greatly overstating the effect of pay for performance. This arises as r only affects the marginal cost of incentive pay.

Measurement Error Worse performance evaluation (σ^2 higher) also renders pay for performance more expensive. However, unlike the previous case, willingness to exert effort for career concerns reasons simultaneously falls such that $\frac{\zeta_\beta}{\zeta_{1-\tau}}$ *increases* in (11). This arises because worse monitoring increases the relative importance of incentive pay despite being more costly.[13]

Hence the effect of agency distortions depends on how it changes the relative weights of various forms of incentive provision.

3.2 Distorted Performance Measures

Those measures for which incentive pay is feasible may differ even on the margin from output produced. A concern of the agency literature is that performance pay distorts actions towards those that are more easily measurable. To this end, consider a case where there are two efforts e_1 and e_2, with costs $C(e_1, e_2)$ where $C_i > 0, C_{ii} > 0$, $C_{ij} > 0$, and $C_1(0, e_2) = C_2(e_1, 0) = 0$.

Let true output be $y = e_1 + e_2 + a + \epsilon$, with the social optimum given by $C_i = 1$ with second order condition $C_{11}C_{22} - C_{12}^2 \geq 0$. This output can be observed but cannot be contracted upon. Instead, incentive contracts can only be based on a distorted measure

$$d = t_1 e_1 + t_2 e_2 \tag{12}$$

[13]It should not be assumed that this is true for all monitoring errors. For example, take the case of σ_a^2 rising. Then career incentives are strengthened (as less is known about the worker) but the marginal cost of pay for performance rises. Both effects reduce the relative weight of pay for performance and increase the difference between $\zeta_{1-\tau}$ and ζ_β.

for $t_i \geq 0$, where $t_1 \neq t_2$.[14] The multitasking problem is that pay for performance distorts towards the effort with the highest t_i. (So for example it could be that one action is unmeasured $t_i = 0$ while the other is measured perfectly, $t_j = 1$.) Career concerns incentives are unchanged as y can be observed, but the agent's marginal incentive pay reward from increasing e_i is $t_i\beta$. The agent chooses therefore effort levels given by $C_1 = (\beta t_1 + \gamma s)(1-\tau)$ and $C_2 = (\beta t_2 + \gamma s)(1-\tau)$.

The tax elasticity overestimated the effect of incentive pay in the baseline model above because tax reductions offer more reasons to exert effort than incentive pay. Here tax reductions increase output for career concerns reasons if and only if $C_{11} + C_{22} \geq 2C_{12}$ (this is the analog to $C'' > 0$ in the single effort case). If this condition holds, we show in the Appendix that $\frac{\zeta_\beta}{\zeta_{1-\tau}} \leq 1$. As a result, the insights above are robust to allowing distorted performance measures in this way.

4 Empirical Testing of Optimal Contacts

The second interest of the paper is in understanding the weak empirical support for the predictions of the most basic theory. One implication of the agency section above is that comparative statics depend on how parameters affect the *relative* importance of various sources of incentives. Yet most empirical work estimating optimal contracts only considers a single reason for exerting effort. We now show how even the simplest comparative statics become difficult to sign when workers have alternative reasons for exerting effort. As taxes are irrelevant here, they are ignored.

Of importance here is that factors that affect pay for performance also affect the marginal return to exerting effort for other reasons. To flesh this out further, we add one more source of incentives, namely promotion. Following Lazear and Rosen, 1981, assume that the agent competes for a promotion with another ex ante identical agent. The competing agent operates a similar technology to (1), and all error terms are independent between the two agents. The agent with the highest observed output is promoted and receives an additional exogenous reward Δ for doing so.[15]

If I_p is an indicator variable equal to 1 if the agent is promoted, then the agent's income is $w + I_p\Delta + \gamma E(a)$. Assume a symmetric pure strategy equilibrium. Then (see Appendix)

[14] These results also hold when there are separate distorted measures $d_i = t_i e_i$ for each e_i, so each has its own incentive pay intensity β_i.

[15] This is treated as exogenous as a way of identifying the effect of pay for performance holding all other incentives constant.

the choice of effort is then given by

$$C'(e^*) = \beta + \gamma s + \frac{\Delta}{2\sqrt{\pi(\sigma_a^2 + \sigma^2)}}. \tag{13}$$

The final term $\frac{\Delta}{2\sqrt{\pi(\sigma_a^2+\sigma^2)}}$ is the marginal change in expected promotion rewards by exerting more effort. Not surprisingly, this depends on the ability to monitor performance. The optimal contract is given by

$$\beta^* = \frac{1 - \frac{\Delta}{2\sqrt{\pi(\sigma_a^2+\sigma^2)}}}{1 + r(\sigma^2 + \sigma_a^2)C''} - \gamma s. \tag{14}$$

Much empirical work in agency theory tests the idea that worse performance measurement constrains incentive pay. Theoretically this is because noisier measures increase the worker's risk premium (the tradeoff of risk and incentives): in the absence of any other incentives, $\frac{d\beta^*}{d\sigma^2} < 0$ for $r > 0$. However, empirical support for this idea has been weak (Prendergast, 2002).

These tests are identified by the assumption that worse measurement increases the marginal cost of providing incentives but does not affect its benefit. This is unlikely to be true with other incentives. Consider the two other incentives described above. Measurement error reduces career incentives (s is decreasing in σ^2), and the marginal return to promotion effort also declines. This implies that the marginal *benefit* to pay for performance simultaneously increases. As a result,

$$\frac{d\beta^*}{d\sigma^2} = \frac{\frac{\Delta}{2\sqrt{\pi}(\sigma_a^2+\sigma^2)^{\frac{3}{2}}} - rC''\beta^*}{1 + r(\sigma^2 + \sigma_a^2)C''} + \frac{\gamma s}{\sigma_a^2 + \sigma^2}, \tag{15}$$

which not surprisingly cannot be signed. A simple way of understanding the ambiguity can be seen by varying the degree of risk aversion in Proposition 1.

Proposition 1 *Optimal pay for performance is increasing in σ^2 for low levels of risk aversion (r), and decreasing for sufficiently high levels of risk aversion.*

An assumption often made is that the wealthy exhibit a greater tolerance for risk than those with fewer resources. As a result, this would suggest that we are likely to find a negative tradeoff between risk and incentives for less well paid occupations, but the reverse for highly paid occupations.

4.1 The Returns to Effort Exertion

It is more important that effort be exerted in some settings than in others. To take an extreme example, it is likely more important that a CEO of a large company has incentives than does the janitorial staff. Accordingly, consider the case where the marginal return to effort for the agent is no longer unity in (1), but instead θ, where $y = \theta e + a + \epsilon$. In the absence of other sources of incentives ($\sigma_a^2 = 0$ and $\Delta = 0$), the principal chooses β to maximize expected surplus subject to $C'(e) = \theta\beta$. The optimal choice of incentives is then computed to be

$$\beta^* = \frac{1}{1 + r\sigma^2 \frac{C''}{\theta^2}}, \tag{16}$$

which is increasing in θ. This captures the intuitive logic that when effort exertion is more important, incentives rise.

Now consider career concerns. When the returns to effort rise, so also are the returns to perceived ability likely to rise. To use the example above, if the value of exerting effort by a CEO is high, so also is the return to his talent likely to be. Hence, the marginal return to ability is also likely to increase in θ. This renders the comparative statics above more complicated because while θ increases the desire by the principal for more effort, the agent may simultaneously be more willing to exert effort without incentive pay.

Accordingly, assume that the agent has career concern incentives ($s > 0$), where the future value of perceived ability is given by $\gamma(\theta)$, where $\gamma'(\theta) > 0$. Hence when effort is valuable, so also is ability. Promotion incentive are ignored. In this case, the effort exerted by the agent is given by $C'(e) = \theta\beta + \gamma(\theta)s$, and the optimal contract is given by

$$\beta^* = \frac{1}{1 + r(\sigma^2 + \sigma_a^2)\frac{C''}{\theta^2}} - \frac{s\gamma(\theta)}{\theta}, \tag{17}$$

with

$$\frac{d\beta^*}{d\theta} = \frac{\frac{2\beta^* r(\sigma^2+\sigma_a^2)C''}{\theta^3} - \left(\frac{\gamma'(\theta)\theta - \gamma(\theta)}{\theta^2}\right)}{1 + r(\sigma^2 + \sigma_a^2)\frac{C''}{\theta^2}}. \tag{18}$$

This can no longer be unambiguously signed as the parameter that increases the desire for more effort simultaneously increases agent willingness to exert it. This ambiguity can be most easily seen by considering the case of risk neutrality.

Proposition 2 *Let $r = 0$. Then optimal pay for performance is decreasing (increasing) in θ if $\gamma'' > 0 (< 0)$.*

In words, with risk neutrality, incentives fall in θ with convex returns to ability. Once again alternative sources of incentives render comparative statics difficult to sign, and may be a reason for the weak empirical performance of these tests.

Given that workers typically have many sources of incentives, what kind of comparative statics are likely to be robust? The most likely avenue for successful testing is where costs change for only *one* form of oversight, as in Bertrand and Mullainathan, 2003, who examined the effect of a change in hostile takeover legislation on the compensation of CEOs. As these changes only affect one form of oversight, they can be used as reliable sources of identification.

5 Conclusion

We have argued here for using labor supply data to estimate the effect of incentive pay on performance. It is worthwhile to conclude with three limitations. First, this approach is most suited to estimating short run effects. Over the long run, differences between tax and incentive pay effects could arise through decisions made by firms. For example, higher labor costs induced by increased taxes could affect investment decisions by firms. No claims are being made about long run effects.[16] Second, it is important to remember that the elasticity estimates address local changes in pay for performance from their current levels. So, for a worker currently on no pay for performance, this exercise measures the impact of moving to a small amount. No statements are being made large discrete changes in pay for performance.

Third, we have ignored the selection effects of incentive pay. The effect of pay for performance goes beyond simply effort exertion: it also attracts better quality workers (Lazear, 2000). Such selection effects, where workers potentially change employers, are included in the taxable income elasticities cited above. However, it is difficult to claim that "market effects" can be ignored here. Specifically, when a firm unilaterally increases incentive pay, that firm is more likely to attract better talent than in the case when taxes are reduced for all firms. This is a potentially important difference between the two cases. However, it is not clear which estimate is most desired. Current studies such as Lazear, 2000, address selection effects by considering changed outputs and profits for the changing firm. Yet if our interest is in the change in surplus from a firm increasing its pay for performance, only considering the effects on the firm that changes compensation likely overstates changes in surplus as other firms lose their better workers to that firm. By contrast, the tax estimates consider the effects on all parties. As a result, it is not clear which measure is preferred.

Another selection issue which has been ignored concerns labor participation. Labor supply hours responses on the extensive margin - where workers choose to enter or leave the labor force based on tax rates - are typically larger than those cited above on the

[16] There is one potential short run response by firms that could confound the exercise here, where a firm responds to a tax change by changing pay for performance, so the empirical estimates confound both effects. However, note from (10) that the optimal contract is independent of τ so this is unlikely to be a major concern.

intensive margin and are in the -0.4 range. Although such participation issues are far from the emphasis of the pay for performance literature, they could conceivably be important, particularly for widespread adoption of incentive pay plans.

To sum up, obtaining widespread empirical estimates of the effect of incentive pay has been understandably difficult. This structural approach attempts to do so, by relying on the fact that public finance has long been interested in mapping the marginal cost of labor inputs through tax changes. By doing so, we hope to extend our understanding of incentive pay beyond the small number of occupations for which information can be directly estimated.[17] However, we are limited in what we can say as workers have other sources of incentives. This issue also pervades the the second part of the paper such that testing optimal contracts becomes difficult. As most workers have some other source of incentives, this renders empirical testing far more complex than is usually considered.

[17]There is much interest in the effects of pay for performance on CEOs. Unfortunately, this approach offers little to cast light on this, as these groups are largely excluded here, because they tend to be those most likely to move income in ways that avoid tax.

References

Aarbu, Karl and Thor Thoreson, 2001, "Income Responses to Tax Changes -Evidence from the Norwegian Tax Reform", *National Tax Journal*, 54(2): 319-334.

mimeo.

Ariely, D., U. Gneezy, G. Loewenstein, and N. Mazar (2009), "Large Mistakes and Big Mistakes," *Review of Economic Studies*, 2009.

Oriana Bandiera, Iwan Barankay, and Imran Rasul, 2005, "Social Preferences and the Response to Incentives: Evidence from Personnel Data", *The Quarterly Journal of Economics,* vol. 120(3), pages 917-962.

Bertrand, Marianne and Sendhil Mullainathan (2003), "Enjoying the Quiet Life? Corporate Governance and Managerial Preferences", *Journal of Political Economy*, 111, 1043-75.

Blundell, Richard and Thomas MaCurdy, 1999, "Labor Supply: A Review of Alternative Approaches." In O. Ashenfelter, D. Card, ed., *Handbook of Labor Economics*, Amsterdam: North-Holland.

Chetty, Raj (2012), "Bounds on Elasticities with Optimization Frictions: A Synthesis of Micro and Macro Evidence on Labor Supply", *Econometrica*, 80(3), 969-1018.

Chetty, Raj, Adam Looney, and Kory Kroft (2008), "Incidence and Taxation: Theory and Evidence", mimeo.

Frey, Bruno, and Reto Jegen, "Motivation Crowding Theory: A Survey of Empirical Evidence," *Journal of Economic Surveys*, 2001, Vol. 15 (5), 589-611.

Gibbons, Robert and Kevin M. Murphy, "Optimal Incentive Contracts in the Presence of Career Concerns", *Journal of Political Economy*, 1992, 100(3), 468-505 .

Giertz, Seth. 2007, "The Elasticity of Taxable Income over the 1980s and 1990s",*National Tax Journal*, 60(4): 743-768.

Gruber, Jonathan, 1997, "The Incidence of Payroll Taxation: Evidence from Chile", *Journal of Labor Economics*, 15: S72-S101.

Gruber, Jonathan and Emmanuel Saez, 2002, "The Elasticity of Taxable Income: Evidence and Implications", *Journal of Public Economics*, 84: 1-32.

Hansson, Asa, 2007, "Taxpayers' responsiveness to tax rate changes and implications for the cost of taxation in Sweden", *International Tax and Public Finance*, 14(5): 563-582.

Holmstrom, Bengt and Paul Milgrom (1992), "Multi-Task Principal Agent Analyses: Linear Contracts, Asset Ownership and Job Design", *Journal of Law, Economics, and Organization*, 7, 24-52.

Holmstrom, Bengt, "Managerial Incentive Problems: A Dynamic Perspective", *Review of Economic Studies*, 1999.

Keane, Michael (2010), "Labor Supply and Taxes: A Survey", mimeo, University of Technology, Sydney.

Kopczuk, Wojciech, 2005, "Tax Bases, Tax Rates and the Elasticity of Reported Income." *Journal of Public Economics*, 89(11-12): 2093-2119.

Kleven, Henrik and Esben Anton Schultz, 2009, "Identifying the Elasticity of Reported Income Using a Large Panel of Danish Tax Payers", work in progress.

Kruse, Douglas (1993), "Does Profit Sharing Affect Productivity?", mimeo, Rutgers University.

Lacetera, Nicola, Mario Macis, and Robert Slomin (2009), "Will There Be Blood? Incentives and Substitution Effects in Pro-Social Behavior", IZA Working Paper.

Lazear, Edward P, (2000), "Performance Pay and Productivity", *American Economic Review*, 90(5): 13461361.

Lazear, Edward and Sherwin Rosen (1981), "Rank Order Tournaments as Optimal Labor Contracts", *Journal of Political Economy*, 89, 841-64.

Lazear, E. and Shaw, K. (2009), "Wage Structure, Raises and Mobility: An Introduction to International Comparisons of the Structure of Wages Within and Across Firms" in *The Structure of Wages: An International Comparison*, Edward Lazear and Kathryn Shaw (editors), University of Chicago Press.

Lemieux, T., MacLeod, Bentley, W. and Parent, D. (2009), "Performance Pay and Wage Inequality", *Quarterly Journal of Economics*, 124 (1): pp. 149.

MacLeod, W. Bentley and Daniel Parent (1999), "Job Characteristics and the Form of Compensation", in Solomon Polacheck, ed., *Research in Labor Economics*, Vol. 18, pp. 177242. Stamford, Conn.: JAI.

Mehrotra A, Damberg C, Sorbero M, and Teleki S., 2010, "Pay-for-Performance in the Hospital Setting: What is the State of the Evidence?", *American Journal of Medical Quality*, 2009

Mellstrom, C. and Johannesson, M., 2008, "Crowding Out in Blood Donation: Was Titmuss Right? *Journal of the European Economic Association*, 6, 4, 845-863.

Neal, Derek (2011), "The Design of Performance Pay in Education", NBER Working Paper, 16710.

Piketty, T. and E. Saez., 2007, "How Progressive is the U.S. Federal Tax System? A Historical and International Perspective", *Journal of the Economic Perspectives*, 21, 1, 2-24.

Prendergast, Canice (2002), "The Tenuous Tradeoff of Risk and Incentives", *Journal of Political Economy*, 110(5), 1071-1102.

Saez, Emmanuel, 2003, "The Effect of Marginal Tax Rates on Income: A Panel Study of 'Bracket Creep'", *Journal of Public Economics*, 87: 1231-1258.

Saez, Emmanuel, 2009, "Do Tax Filers Bunch at Kink Points? Evidence, Elasticity Estimation, and Salience Effects," Working Paper University of California at Berkeley, Berkeley, CA.

Saez, Emmanuel, 2012, "Labor Supply Responses to Taxes and Transfers", Berkeley, CA.

Saez, Emmanuel, Joel Slemrod, and Seth Giertz, 2009, "The Elasticity of Taxable Income With Respect to Marginal Tax Rates: A Critical Review'", NBER Working Paper 15012.

Selen, Jan. 2002, "Taxable income response to tax changes: Evidence from the 1990/91 Swedish Tax Reform", FIEF Working Paper No. 177.

Sillamaa, Mary Anne and Michael Veall, 2001, "The Effect of Marginal Tax Rates on Taxable Income: a Panel Study of the 1988 Tax Flattening in Canada," *Journal of Public Economics*, 80(3): 341-356.

Appendix:

The Choice of Effort and the Optimal Contract The firm is competitive, and there are no constraints on transfers, so the firm chooses the contract (β_0, β) to maximize the agent's utility. The worker's utility with a given realization is $-exp\{-r[(1-\tau)[\beta_0 + (\beta + \gamma s)(a + e_i^* + \epsilon)] - C(e_i^*)]\}$, with expected utility

$$-exp\{-r[(1-\tau)[\beta_0 + (\beta + \gamma s)e_i^*] - C(e_i^*)]\}.exp\{\frac{1}{2}r^2(1-\tau)^2(\beta+\gamma s)^2(\sigma_a^2 + \sigma^2)\}. \quad (19)$$

Differentiating this with respect to e_i^* yields $C'(e_i^*) = (1-\tau)[\beta + \gamma s]$. Substituting this into (19) and differentiating with respect to β yields (10).

Distorted Performance Measures Following exactly the same steps as above, it is simple to show that $C_1 = (\beta t_1 + \gamma s)(1-\tau)$ and $C_2 = (\beta t_2 + \gamma s)(1-\tau)$. Then $\frac{dy}{d\beta} = \frac{(1-\tau)t_1 - \frac{C_{12}}{C_{22}}(t_2(1-\tau))}{C_{11} - \frac{C_{12}^2}{C_{22}}} + \frac{(1-\tau)t_2 - \frac{C_{12}}{C_{11}}(t_1(1-\tau))}{C_{22} - \frac{C_{12}^2}{C_{11}}}$ and $\frac{dy}{d(1-\tau)} = \frac{\beta t_1 + \gamma s - \frac{C_{12}}{C_{22}}(\beta t_2 + \gamma s)}{C_{11} - \frac{C_{12}^2}{C_{22}}} + \frac{\beta t_2 + \gamma s - \frac{C_{12}}{C_{11}}(\beta t_1 + \gamma s)}{C_{22} - \frac{C_{12}^2}{C_{11}}}$. As a result,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \left(\frac{\beta(t_1 - \frac{C_{12}}{C_{22}}t_2)}{C_{11} - \frac{C_{12}^2}{C_{22}}} + \frac{\beta(t_2 - \frac{C_{12}}{C_{11}}t_1)}{C_{22} - \frac{C_{12}^2}{C_{11}}}\right) \Big/ \left(\frac{\beta(t_1 - \frac{C_{12}}{C_{22}}t_2) + \gamma s(1 - \frac{C_{12}}{C_{22}})}{C_{11} - \frac{C_{12}^2}{C_{22}}} + \frac{\beta(t_2 - \frac{C_{12}}{C_{22}}t_1) + \gamma s(1 - \frac{C_{12}}{C_{11}})}{C_{22} - \frac{C_{12}^2}{C_{11}}}\right). \quad (20)$$

Consider the γs terms on the denominator. These can be simplified to $\gamma s(\frac{C_{11}+C_{22}-2C_{12}}{C_{11}C_{22}-C_{12}^2})$ which is non-negative if $C_{11} + C_{22} \geq 2C_{12}$. If the γs terms are non-negative, then $\frac{\zeta_\beta}{\zeta_{1-\tau}} \leq 1$ as above.

Adding tournament incentives Two agents, 1 and 2, are employed and let subscript i refer to agent i. The output produced by agent i is given by $y_i = e_i + a_i + \epsilon_i$, where a_i is the agent's ability, and ϵ_i is his error term. If I_{pi} is an indicator variable equal to 1 if agent i is promoted, then this income is $w_i + I_{pi}\Delta + \gamma E(a_i)$. The two agents compete for promotion, and the winner has the highest y_i. In a pure strategy equilibrium, the probability of winning the tournament is given by $\Phi(e_i - \tilde{e}_j)$, where Φ is a Normal distribution with mean 0 and variance $2(\sigma_a^2 + \sigma^2)$, and $\tilde{e}_j$ is the expected equilibrium effort of the other competitor. Let ϕ be the density of this function. The objective of the agent is to maximize $E\{-exp\{-r[\beta_0 + \beta y_i + \gamma s(y_i - E(e_i)) + \Delta\Phi(e_i - \tilde{e}_j) - C(e_i)]\}\}$, where $\tilde{e}_j$ is the equilibrium choice of effort in a symmetric pure strategy equilibrium. This is equal to

$$-exp\{-r[([\beta_0 + (\beta+\gamma s)e_i^* + \Delta\Phi(e_i^* - \tilde{e}_j)] - C(e_i^*)]\}.exp\{\frac{1}{2}r^2((\beta+\gamma s)^2(\sigma_a^2+\sigma^2) + \frac{\Delta^2}{2})\}. \quad (21)$$

Differentiation yields (13) when imposing the condition that $e_i^* = \tilde{e}_j$ and $\phi(0) = \frac{1}{2\sqrt{\pi(\sigma_a^2+\sigma^2)}}$. The optimal contract is then derived identically to above yielding (14).

Changing Returns to Effort With θ as the marginal return to ability, the firm chooses the contract (β_0, β) to maximize the agent's utility. The worker's utility with a given realization is $-exp\{-r[(1-\tau)[\beta_0 + (\theta\beta + \gamma(\theta)s)(a + e_i^* + \epsilon)] - C(e_i^*)\}$, with expected utility

$$-exp\{-r[(1-\tau)[\beta_0+(\theta\beta+\gamma(\theta)s)e_i^*]-C(e_i^*)\}.exp\{\frac{1}{2}r^2(1-\tau)^2(\theta\beta+\gamma(\theta)s)^2(\sigma_a^2+\sigma^2)\}. \quad (22)$$

Optimal effort is given by $C'(e_i^*) = (1-\tau)[\theta\beta+\gamma(\theta)s]$. Substituting and differentiating with respect to β yields (17), and differentiating that with respect to θ yield (18).