The Empirical Content of Pay-for-Performance

Canice Prendergast

University of Chicago Booth School of Business

canice.prendergast@chicagobooth.edu

Abstract

Empirical evidence on the effect of pay for performance on output has been scarce. We propose that worker responses to marginal pay for performance changes can be related to their response to a measure of taxes. Using this approach, we suggest a short run elasticity of output with respect to incentive pay for high earners in the United States of 0.25 or lower, and it is difficult to rule out very low responsiveness.

1 Introduction

Pay for performance has played a prominent role in the economics of organizations. It has been both lauded as the solution to misaligned incentives and criticized as a source of dysfunctional behavior. The enormous literature on this subject is typically premised on the assumption that pay for performance matters for performance. Surprisingly, we know relatively little about whether this is empirically true despite the literature reaching its fourth decade.

There are a number of well known studies that address how incentive pay affects performance using data on individual firms. These exercises are by necessity narrowly focused as they require both personnel data with performance measures, and a 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 existing studies.[1]Consequently, little can be said about how the average worker would respond to changed incentives.

This paper contributes to this debate by offering a more widespread estimate of the response of output to performance pay, by addressing the short run impact of a firm marginally changing performance pay from its current level. To do so, we use the empirical large literature on labor supply. Under the assumption that workers have some reason to exert non-contractible efforts, the contribution of the paper is to illustrate a relationship between the elasticity of output with respect to pay for performance and the elasticity of output with respect to net marginal income (one minus the tax rate), for which there is considerable empirical evidence.

We claim that, for two separate reasons, the response of output to net marginal income overstates the response to performance pay. Both relate to the fact that firms have ways of inducing performance beyond formal pay for performance. First, some worker inputs can be easily observed (attendance, for example). Second, even if inputs cannot be directly observed, there are other ways than formal performance pay to induce non-contractible effort. For example, workers may be willing to exert effort to gain promotion with their current employer, or to attain a better paid job elsewhere. We begin by considering the case where all worker inputs are non-contractible, as it typically assumed in agency settings. We show in that case that the ratio of the tax elasticity to the performance pay elasticity is the fraction of all monetary incentives that pay for performance comprises. As workers have many possible sources of incentives, tax responses typically overstate the response to pay for performance. These measures are equal only when incentive pay is the sole monetary reason to exert effort.

While agency theory is typically based on the assumption that inputs to production cannot be directly contracted upon, it is clear many inputs provided by workers can be easily observed: for example, it is easy to see if someone showed up to work. The second case we consider is where some inputs can be specified without a performance pay contract. Once again, tax responses exceed performance pay responses even if performance pay is the only reason to exert non-contractible effort. This arises because tax changes induce changes in the contractible components of production in a way that does not arise when performance pay changes. As one example, reductions in taxes cause employers to want to increase hours worked as the marginal returns to those activities increase. By contrast, changing performance pay has no such direct effect on hours worked, and so the responsiveness to taxes on output exceeds the responsiveness of performance pay.

This mapping allows us to use the tax literature to inform the impact of marginal incentive pay changes for a broad group of workers. The tax literature primarily focuses on short run responses to tax changes, typically for the top 10% of earners.

Our inferences on the impact of performance pay are also restricted to short run responses for this group. Using existing studies on the relationship between tax rates and taxable income, we suggest an elasticity of output with respect to pay for performance for those high earners of 0.25 or lower, and it is hard to rule out very low responsiveness. This result, which require two additional assumptions - competitive labor markets and the absence of income effects - suggest that for a broad class of workers, the impact of additional incentive pay may be limited. As such, this work also provides a reason why pay for performance is relatively rare, as the marginal returns may be low.[2]

The conclusion drawn from considering the empirical literature on taxes is that, at least over the short run, the responsiveness of worker effort to pay seems quite small. It is important to note that this should be seen not as the primary contribution of the paper, but rather how the methodology can be used to make inferences about short run responses. This conclusion could be overturned if, for example, future studies could find large long run effects of tax changes on incomes. Instead, the paper's primary contribution is to identify the conditions under which tax responses can help to understand the impact of incentive pay.

2 A Model

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

$$y = e, \tag{1}$$

where e is effort exerted. The cost of effort for the agent is given by $C(e)$, where $C'(e) > 0, C''(e) > 0$, and $C'(0) = 0$. We begin by considering the case where all effort is unobserved. by the firm.

The firm potentially offers pay for performance. The agent is paid a (pre-tax) wage w by the firm equal to

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

where $\beta \geq 0$. (Linearity of compensation does not matter for the results below: it simply provides a parameterization for comparative statics below.)

The agent has some other monetary reason for exerting effort beyond pay for performance. A simple case is where future compensation depends on current effort. (A career concerns interpretation is elaborated upon below.) Consider a case where the agent's expected compensation (including future benefits) is $w + m(e)$, where $m' \geq 0$, $m'' \leq 0$. m could be a function of y but need not.

The agent is risk neutral and cares about monetary rewards and effort exerted. 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

$$(1-\tau)[\beta_0 + \beta y + m(e)] - C(e). \tag{3}$$

The outcome of the paper can be easily seen from the objective function. Testing the impact of pay for performance on output is understanding how agents value the tradeoff between money and leisure. Yet tax changes map out a similar role. The marginal value of effort on performance pay is $(1-\tau)\beta$, which naturally implies that there is a likely relationship between the response to taxes and performance pay, as their effects on net incentive pay are identical.

3 The Response to Incentive Pay

A relationship between the effort responses to a measure of taxes and to pay for performance is now shown. To do so, we consider the effect of marginally changing performance pay from current levels. This potentially depends on how pay for performance levels are set. One possibility is that pay for performance is optimally set. However, the results of Lazear, 2000, showed large increases in profits when pay for performance was introduced, suggesting that pay for performance was set at suboptimal levels. Accordingly, we consider both the case of optimally set incentives and where incentives are exogenously set.

Consider the agent's incentives. Optimal effort is characterized by

$$C'(e^*) = (1-\tau)[\beta + m']. \tag{4}$$

Intuitively, effort is increasing in pay for performance and other incentives, and decreasing in tax rates.

Begin by considering the case where β is exogenously set, and is marginally changed.[3]The elasticity of output with respect to incentive pay is given by

$$\zeta_\beta = \frac{\partial y}{\partial \beta}\frac{\beta}{y} = \frac{(1-\tau)\beta}{e^* C''(e^*)}. \tag{5}$$

Now consider the outcome when taxes change, holding β fixed. The elasticity of output with respect to "net marginal income" $1-\tau$ is given by

$$\zeta_{1-\tau} = \frac{\partial y}{\partial (1-\tau)}\frac{(1-\tau)}{y} = \frac{(1-\tau)(\beta + m')}{e^* C''(e^*)}. \tag{6}$$

Hence,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \frac{\beta}{\beta + m'} \leq 1. \tag{7}$$

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. The reason for higher response to $1 - \tau$ is that taxes affect all reasons for exerting effort, while the pay for performance elasticity only captures that one reason for exerting effort. As a result, identifying the effects of a change in taxes on output weakly overestimates the impact of a change in incentive pay. The two measures are equal only if pay for performance is the sole reason for effort exertion.

So far, we have considered incentive pay to be exogenous, which allows us only to bound the performance pay elasticity. We can make more progress when the firm optimally chooses incentive pay at β^*. With optimally chosen incentives,

$$\zeta_{1-\tau} = \left[\frac{\partial y}{\partial(1-\tau)} + \frac{\partial y}{\partial \beta^*}\frac{\partial \beta^*}{\partial(1-\tau)}\right]\frac{(1-\tau)}{y} \tag{8}$$

Note that optimal pay for performance is $\beta^* = 1 - m'$. Hence the outcome in (7) is not affected by endogenous incentive pay but rather simply evaluated at β^* above. Furthermore, at the optimum

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \beta^* \tag{9}$$

so that a precise estimate for the incentive pay elasticity can be attained from tax data if the level of incentive pay is observed.

3.1 Monitoring Inputs

Here we provide another reason why tax responses may overstate the effect of performance pay: that there are some components of worker performance that are easily observed and do not need incentive pay. For example, in many occupations superiors can directly observe hours worked, and more generally supervisors often directly observe the actions of the subordinates. It is shown here that the ability to observe a component of worker inputs provides an independent reason for tax responses overstating the response to performance pay. The is because when taxes change, firms increase the contractible inputs to performance for a reason that does not arise when performance pay is increased.

To address this, consider a case where the principal can observe some contractible component of performance κ, but cannot observe intensity of effort e, where output is $y = \kappa + e$, and the cost of inputs is $C(\kappa, e)$ where $C_i > 0$, $C_{ii} > 0$ and $C_{e\kappa} > 0$.[4]

The difference here is that the firm can directly contract on κ to maximize $(1 - \tau)\kappa - C(e, \kappa)$ and so satisfies the first order condition $C_\kappa(e, \kappa) = (1 - \tau)$. Consider the case where it does such "input monitoring".[5]As a result, totally differentiating the first order condition yields

$$\frac{d\kappa}{d(1-\tau)} = \frac{1}{C_{\kappa\kappa}} - \frac{C_{e\kappa}}{C_{\kappa\kappa}} \frac{de}{d(1-\tau)} \tag{10}$$

By contrast, the impact of performance pay on κ is indirect, though its impact on effort. Totally differentiating $C_\kappa(e, \kappa) = (1 - \tau)$ yields

$$\frac{d\kappa}{d\beta} = -\frac{C_{e\kappa}}{C_{\kappa\kappa}} \frac{de}{d\beta}. \tag{11}$$

Then as $\frac{dy}{d(1-\tau)} = \frac{de}{d(1-\tau)} + \frac{d\kappa}{d(1-\tau)}$,

$$\frac{dy}{d(1-\tau)} = (1 - \frac{C_{e\kappa}}{C_{\kappa\kappa}}) \frac{de}{d(1-\tau)} + \frac{1}{C_{\kappa\kappa}}, \tag{12}$$

and similarly

$$\frac{dy}{d\beta} = (1 - \frac{C_{e\kappa}}{C_{\kappa\kappa}})\frac{de}{d\beta} < \frac{dy}{d(1-\tau)} \tag{13}$$

if either $C_{\kappa\kappa} > 0$ or $\frac{de}{d\beta} < \frac{de}{d(1-\tau)}$. Once again, the responsiveness of output to taxes exceeds the responsiveness to incentive pay, though here it does not rely on the presence of other sources of effort incentives. For example if $m' = 0$ so the agent has no other reason to exert effort,

$$\frac{dy}{d(1-\tau)} - \frac{dy}{d\beta} = \frac{1}{C_{\kappa\kappa}}, \tag{14}$$

and so the responsiveness to taxes still exceeds that of performance pay.

The reason is that contractible inputs respond to lower taxes for two reasons. First, the surplus from the employment relationship rises and the firm moves up the labor supply curve. Second, lower taxes increase effort, and as increased effort increases the marginal cost of hours, that chokes off some of that increase. By contrast, only the second effect arises when incentive pay increases. Hence, the response to taxes again provides an upper bound on the responsiveness to incentive pay, though here it arises through the endogenous response of the contractible input.

To understand the extent to which tax elasticities overstate the response to incentive pay, one would want empirical estimates of the importance of pay for performance in effort decisions compared to (i) other monetary reasons for exerting effort and (ii) the ability to directly monitor performance though inputs like hours. Unfortunately, no such estimates exist. Yet $\frac{\zeta_\beta}{\zeta_{1-\tau}}$ is increasing in β in (7), and empirical evidence consistently finds low rates of formal pay for performance in firms. As a result, observed tax elasticities may substantially overstate responses to pay for performance.

3.2 Caveats

There are two potential sources of caveats that should be noted. First, the agent's incentives above are determined by net payoffs, so that when considering exerting effort for performance pay reasons, output is discounted by $\beta(1-\tau)$. This requires that two conditions hold. First, as it typically assumed, agents care about their net pay from their labor supply rather than gross pay. Second (and related), workers must perceive changes in tax rates as equivalent in their motivation to changes in incentive pay. Two issues may limit this. First, salience effects must be limited. If the agent notices pay for performance more than he notices taxes (Chetty et al, 2009), the bounds described here may not hold as they care less about the mapping from gross to net pay than they do the effect on increasing gross pay through greater effort. Second, the complexity of the tax code may give rise to confusion in true marginal returns. While this could make perceived marginal tax rates higher or lower than the truth, it does allow the possibility that agents underestimate the impact of tax changes on their pay.[6]

The second important caveat is that the results rely on the agent having some monetary reason for exerting *unobservable* effort. When $\beta + m' = 0$, the agent has no incentive to exert effort. Then if β is changed to be small but positive, the tax elasticities based on no incentives cannot inform what happens when incentives become positive. As a result, an identifying assumption for this approach is that the agent has some monetary reason for exerting effort, if not for performance pay reasons then for the prospect of promotion or moving to a new job.

A relevant example may be where the contractible measure κ is hours worked, in a setting where workers have no additional reason to exert effort. In that setting, the responsiveness of output to $(1-\tau)$ would be $\frac{1}{C_{\kappa\kappa}}$ as only hours change. But knowing something about C_{kk} not tell us anything about the response of workers from going to zero effort incentives to positive effort incentives, which depends on C_{ee}. As a result, care should be taken in using these results to infer the responses of such workers, and no claim is being made for their responsiveness. Instead, the model applies to cases where workers already have reasons to exert non-contractible effort. Apropos of this, the empirical evidence on taxable income, to which we now turn, only concerns the responses of high earners, typically those in the top 10% of the earnings distribution, where such concerns seem less relevant.

4 An Application to Taxable Income

This mapping between the effect of taxes and incentive pay could be used in many ways. For example, it could be used on personnel data, where the effect of tax changes on output could be used to infer the impact of a hypothetical incentive pay change. Alternatively, there is a large literature on the effect of tax changes on labor 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 in say a particular occupation, we now turn to an broad estimate of responsiveness to pay for performance.

Specifically, we use evidence on those who pay taxes and whose labor supply responses have been mapped to address responses to incentive pay. In the United States, this is a relatively small part of the economy, typically the top 10% of earners, and our estimates are specific to that group. Furthermore, these estimates are short run responses, so our inferences must also be restricted to short run effect. Finally, note that the model requires that workers are not simply being monitored on inputs like hours, but rather also are willing to exert non-contractible effort. Focusing on high earners would seem appropriate for this additional reason as they are likely to be doing more than the minimum that their employers can specify through their ability to monitor.

Thus far, we have not used the assumption that the market is competitive. Competition allows us to map from the theoretical results on output (which is not observed in most data sets) 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^*. 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.

Saez et al, 2009, review the literature on taxable income, and this section owes much to that work. Based on their overview, they conclude that 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.

There are two components of tax avoidance that are relevant. First, one strand of the literature 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 (see Saez et al, 2009, for a discussion). This effect strengthens the overestimate of taxes for pay for performance, because for example a tax reduction would increase taxable income not because output rises but because the fraction of all income that is taxable would also increase. The second form of avoidance is not through changed composition but various strategies to make a given output less taxable. 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. 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.[7]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 smaller than -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 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.[8]

On the basis of this, -0.2 seems a high 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 a generous 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 our 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. 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, which suggests a limited role for incentive pay in increasing output. Finally, we noted that the use of pay for performance in firms is relatively rare. This paper could offer a justification for this: it could simply be that on the margin it has little impact on output. However, while this exercise suggests that the short run response of worker effort to pay seems quite small, it should be pointed out that this is not the primary contribution of the paper. For example, future studies may illustrate large long run effects of tax changes on incomes, which would also potentially apply to the effect of incentive pay. Instead, the paper's contribution is primarily to understand the circumstances under which a mapping can be made from tax responses to those for incentive pay.

4.1 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.[9]

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 (though see Bingley and Lanot, 2002, suggesting shared incidence).

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. (Results on the extensive margin are discussed below.) 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. By considering risk neutral agents, such effects can be ignored. More generally, consider the effect of a tax increase. With incidence on the agent, his income (holding effort constant) falls. As pay for performance has no such income effect,[10]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.

Another potentially important issue is adjustment costs (Chetty et al, 2011). A recent literature addresses difficulties that workers face in changing their hours response to tax changes. For example, a worker who is in a position that involves 40 hours a work on the job may not be easily able to change her hours to 45. A few responses to this issue seem relevant. First, such constraints are also likely to be relevant for how workers respond to pay for performance changes in the short run. Second, even if hours of work are fixed, tax changes should give a reason for workers to change the intensity of their efforts in a way that is also informative of their responses to pay for performance change. Yet despite this, much of the literature is concerned with the possibility that short run responses to tax changes are likely to be low for the reason that there are constraints to behavioral changes. Perhaps over the long run such changes arise through changing employers. No claims are being made here about long run responses to pay for performance: in the same way as the tax literature estimates short run responses, so also is this paper addressing short run responses to incentive pay.[11]

The final caveat was mentioned above: these results may tell us little about the use of pay for performance for workers who have no reason to exert non-contractible efforts. This is why we focus on the top 10% of earners, as they are likely doing more than simply showing up for work and clocking off when five o' clock comes.

5 Agency Concerns

A possible criticism of this work is that agency theory is largely concerned with situations where the provision of incentives is costly, yet the model above allows true output to be contracted upon with risk neutral agents. As such, it ignores the central concerns of that literature, namely risk aversion and distorted performance measures. The purpose of this section is to illustrate the robustness of the insights above to these concerns. We do so in a setting where we additionally flesh out the "other" incentives as these may also potentially interact with parameters of interest. Accordingly, we begin by endogenizing the other source of incentives $m(e)$ by providing a career concerns interpretation. We then use that model to consider the case of a risk averse agent. Finally, we consider the other primary concern of agency theory namely, the impact of distorted performance measures. Subject to one caveat concerning a risk premium, the results above continue to hold.

5.1 Other Incentives

Here we provide a more developed model of the agent's other incentives by considering the impact of career concerns. Consider a case where output is

$$y = e + a + \epsilon, \tag{15}$$

where a is the ability of the agent. 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 distribution of ϵ is assumed to be Normal with mean 0 and variance σ^2. The two error terms are uncorrelated.

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.[12]Note that the same performance measure y is used for both sources of incentives.

The objective of the agent is then to maximize $E\{(1-\tau)[\beta_0 + \beta y + \gamma E(a|y)] - C(e)\}$.[13]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

$$E\{(1-\tau)[\beta_0 + \beta y + \gamma s(y - E(e))] - C(e)\}. \tag{16}$$

Optimal effort is characterized by $C'(e^*) = (1-\tau)[\beta+\gamma s]$. The elasticity of output with respect to incentive pay is given by $\zeta_\beta = \frac{\beta(1-\tau)}{y(e^*)C''(e^*)}$, while the elasticity of output with respect to "net marginal income" $1-\tau$ is given by $\zeta_{1-\tau} = \frac{(1-\tau)(\beta+\gamma s)}{y(e^*)C''(e^*)}$. Hence

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \frac{\beta}{\beta+\gamma s} \leq 1, \tag{17}$$

and so the results continue to hold in this more developed model of incentives.

5.2 Risk Aversion

One of the central themes of agency theory is that risk aversion limits the use of incentive pay. This section addresses the robustness of the insights to allowing risk averse agents. We do so in the career concerns setting where closed form solutions are possible. We ignore income effects by considering exponential utility: if the agent's income is I, his utility is

$$U(I,e) = -exp(-r(I - C(e))). \tag{18}$$

With the components of pay outlined above, the objective of the agent is to maximize

$$E\{-exp\{-r[(1-\tau)[\beta_0 + \beta y + \gamma s(y - E(e))] - C(e)]\}\} \tag{19}$$

As above, effort is given by $C'(e^*) = (1-\tau)[\beta + \gamma s]$.

In the previous section, we considered the case of both exogenous β and when incentives are optimally set. First, for any exogenous β - meaning incentive pay that does not change with the tax rate - the bounds described above are unchanged, where $\frac{\zeta_\beta}{\zeta_{1-\tau}} = \frac{\beta}{\beta+\gamma s} \leq 1$. Hence risk aversion does not change the outcome when piece rates are fixed.

The only caveat is with endogenously set incentive pay. The optimal choice of incentives (following Gibbons and Murphy, 1992, and shown in the Appendix) is given by

$$\beta^* = \frac{1}{1 + r(\sigma^2 + \sigma_a^2)C''} - \gamma s. \tag{20}$$

This does not depend directly on tax rates, but could indirectly via the C'' term. When incentives are optimally set

$$\zeta_{1-\tau} = \left[\frac{\partial y}{\partial(1-\tau)} + \frac{\partial y}{\partial\beta^*}\frac{\partial\beta^*}{\partial(1-\tau)}\right]\frac{(1-\tau)}{y}$$
$$= \frac{(1-\tau)(\beta^* + \gamma s)}{yC''}\left[1 - \frac{r(\sigma^2 + \sigma_a^2)C'''}{1 + r(\sigma^2 + \sigma_a^2)C''}\right]. \tag{21}$$

As a result,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \frac{\beta^*}{(\beta^* + \gamma s)(1 - \frac{r(\sigma^2+\sigma_a^2)C'''}{1+r(\sigma^2+\sigma_a^2)C''})}. \tag{22}$$

This differs only from the cases of risk neutrality (or risk aversion with exogenous β) only through the final term in the denominator, which measures the endogenous response of incentive pay to tax rates. With risk neutrality, this term is zero. The term is also zero with quadratic costs of effort ($C''' = 0$), in which case $\frac{\zeta_\beta}{\zeta_{1-\tau}}$ is identical to the risk neutrality case in (7), though evaluated at β^* in (20).[14]

For risk averse agents and non-zero third derivatives of the cost function, the outcome changes from the risk neutral benchmark. If $C''' < 0$, then the tax elasticity remains lower than the incentive elasticity, as before, and is strengthened by the response of incentive pay to taxes. However, if $C''' > 0$, then the right hand side of (22) could exceed 1, as a result of the endogenous response of incentives to taxes. If this effect is sufficiently strong, then using tax changes as an upper bound on incentive pay changes may no longer hold.

This exception to the bounding result above relies on a number of conditions simultaneously holding to affect the taxable income exercise above. First, workers must be risk averse. Second, it requires that the third derivative of the cost function is sufficiently important to result in incentives changing in countervailing ways to a tax change. Third, it must be that workers do not have significant alternative reasons to exert effort. Finally, these incentive pay changes must occur by the time the tax estimates are attained. While these conditions simultaneously holding seem unlikely in most settings, this section does stress that some care should be taken in assuming the bound holds.

5.3 Distorted Performance Measures

The second major concern of the agency literature is when the measure on which the agent is paid does not accord with the principal's objective (Holmstrom and Milgrom, 1991). This section addresses the robustness of the insights when where performance measures are distorted on the margin.

To allow for the distortion of efforts across activities, consider the risk neutral career concerns case where there are two non-contractible 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 $y = e_1 + e_2 + a + \epsilon$. This can be observed but cannot be contracted upon. Efficiency requires that the marginal costs of each effort are equated at $C_i = 1$ with second order condition $C_{11}C_{22} - C_{12}^2 \geq 0$. However, the problem is that incentive contracts can only be based on a distorted measure

$$d = t_1 e_1 + t_2 e_2 \tag{23}$$

for $t_i \geq 0$, where $t_1 \neq t_2$.[15]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 $(1 - \tau)t_i\beta$.

The tax elasticity exceeds the effect of incentive pay in the baseline model because tax reductions offer more reasons to exert effort. 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.

6 Conclusion

We have argued here for a value for considering labor supply data as an avenue to estimating the effect of incentive pay on performance. The arrival rate of new empirical studies on this topic is low, as data requirements are high. Furthermore, it is hard to know the generalizability of these studies to other settings. This paper proposes an avenue to developing a more widespread estimate of the impact of pay for performance for workers who both pay taxes and exert more than minimum effort. Using this methodology, we argued that based on current estimates, the responsiveness of worker inputs to pay seems quite small.

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. Furthermore, long responses by workers are likely to exceed short run responses as adjustment costs by workers could be overcome through, say, job changes, which take time. No claims are being made about long run effects. Second, it is important to remember that the elasticity estimates address local changes in pay for performance from their current levels. 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 when taxes are reduced for all firms. This is a potentially important difference between the two cases.[16] 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 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.

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.

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

Bingley, G. and P. Lanot. 2002. "The Incidence of Income Tax on Wages and Labour Supply", *Journal of Public Economics*, 83, 173-94.

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

Chetty, Raj, John N. Friedman, Tore Olsen, and Luigi Pistaferri. 2011. "Adjustment Costs, Firm Responses, and Micro vs. Macro Labor Supply Elasticities: Evidence from Danish Tax Records", *Quarterly Journal of Economics*, 126(2), 749-804.

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. 2009. "Incidence and Taxation: Theory and Evidence", *American Economic Review*, 99(4), 1145-1177.

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

Gibbons, Robert and Kevin M. Murphy. 1992. "Optimal Incentive Contracts in the Presence of Career Concerns", *Journal of Political Economy*, 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. 1991. "Multi-Task Principal Agent Analyses: Linear Contracts, Asset Ownership and Job Design", *Journal of Law, Economics, and Organization*, 7, 24-52.

Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective", *Review of Economic Studies*, 66(1), 169-182.

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, 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.

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

Risk Aversion: 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 agent'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

$$-exp\{-r[(1-\tau)[\beta_0+(\beta+\gamma s)e_i^*]-C(e_i^*)]\}.exp\{\frac{1}{2}r^2(1-\tau)^2(\beta+\gamma s)^2(\sigma_a^2+\sigma^2)\}. \quad (A1)$$

Differentiating this with respect to e_i^* yields $C'(e_i^*) = (1-\tau)[\beta+\gamma s]$. Substituting this into (A1) and differentiating with respect to β yields (20).

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,

$$\frac{\zeta_\beta}{\zeta_{1-\tau}} = \left(\frac{\beta(t_1 - \frac{C_{12}}{C_{22}}t_2)}{C_{11}-\frac{C_{12}^2}{C_{22}}} + \frac{\beta(t_2 - \frac{C_{12}}{C_{11}}t_1)}{C_{22}-\frac{C_{12}^2}{C_{11}}}\right) \Big/ \left(\frac{\beta(t_1 - \frac{C_{12}}{C_{22}}t_2)+\gamma s(1-\frac{C_{12}}{C_{22}})}{C_{11}-\frac{C_{12}^2}{C_{22}}} + \frac{\beta(t_2 - \frac{C_{12}}{C_{22}}t_1)+\gamma s(1-\frac{C_{12}}{C_{11}})}{C_{22}-\frac{C_{12}^2}{C_{11}}}\right). \quad (A2)$$

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.

Footnotes * I am grateful to the editor, two referees, and seminar participants at the NBER, Columbia University, and Duke University for helpful comments.

1. For example, Lazear, 2000, 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. Mellstrom and Johannson, 2008, find little evidence of blood donation responding to marginal incentives while and Lacetera et al., 2010, find large effects. 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.

2. 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.

3. 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 (4), the optimal choice of effort is independent of the worker's expected equilibrium utility, so that any market effects caused by tax incidence have no effect on effort choices.

4. These needs to be strict substitutes, or else the firm would simply specify κ at the optimal level and allow $e = 0$.

5. We are addressing an increase in β from an exogenously set level. With optimally chosen β using pay for performance and specifying inputs are perfect substitutes as there is no cost to providing incentives through pay for performance here. However, so long as there is any marginal cost to using pay for performance (as addressed in

Section 5), the firm will directly contract on κ to eliminate those distortions.

6. Of course, many pay for performance plans are also quite complicated and so it is unclear which way this effect would bias the mapping.

7. Of course, they also have more flexibility over hours decisions too.

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

9. 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 $-z$.

10. This is because the worker's fixed pay is adjusted to guarantee that he is no better off in utility terms than he was before the pay for performance change. This is not true for tax changes.

11. It is useful to note here that while the distinction between short and long run responses to tax changes has played a central role in empirical work, this distinction has played little role in studies of pay for performance.

12. 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.

13. We are ignoring discounting here, or the possibility that future tax rates differ from current ones.

14. For quadratic costs $C(e)=\frac{ce^2}{2}$, $\frac{\zeta_\beta}{\zeta_{1-\tau}}=1-\gamma s(1+r(\sigma^2+\sigma_a^2)c)\leq 1$.

15. These results also hold when there are separate distorted measures $d_i=t_ie_i$ for

each e_i, so each has its own incentive pay intensity β_i.

16. 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.