Optimal Incentives under Moral Hazard: From Theory to Practice*

George Georgiadis and Michael Powell†

April 10, 2019

Preliminary & Incomplete

Abstract

This paper addresses the following practical question: given an existing incentive contract, what information must a manager acquire to determine how to improve upon that contract? We use a canonical principal-agent framework under moral hazard and assume the principal has productivity data corresponding to some *status quo* contract. Our main result shows that if the principal has *a priori* information about the agent's marginal utility function, and she carries out an experiment in which she perturbs the existing contract, then she can estimate how the agent will respond to a change in his marginal incentives, as well as how the agent's marginal incentives will respond to *any* other perturbed contract. The information provided by such an experiment, therefore, serves as a sufficient statistic for the question of how to locally improve upon the existing contract optimally. The same informational requirements hold, and an analogous sufficient statistic result is obtained, when the principal is restricted to choosing from a lower-dimensional parametric class of contracts; *e.g.*, linear contracts. We also describe the informational requirements for assessing global optimality, where local information like the type described above is insufficient.

*We are grateful to Iwan Barankay, Dan Barron, Hector Chade, Ben Golub, Eddie Lazear, Nathan Seegert, and Jeroen Swinkels, as well as to participants at several seminars and conferences for helpful comments. Finally, we thank Henrique Brasiliense De Castro Pires for excellent research assistance.

†Kellogg School of Management, Northwestern University, Evanston, IL 60208, U.S.A., g-georgiadis@kellogg.northwestern.edu and mike-powell@kellogg.northwestern.edu .

1 Introduction

Research on agency theory has flourished since the seminal works of Mirrlees and Holmström in the 1970s (see, for example, Holmström (2017) and the references within). As a result, we know a lot about optimal contract design; specifically, what are the key trade-offs, formulas that characterize the optimal contract, comparative statics, and so on. Moreover, a growing literature explores the degree to which contracts observed in practice are consistent with theoretically optimal ones (see, for example, Chiappori and Salanié (2003) and the references within). However, we still lack answers to many basic practical questions. For instance, consider a manager who wants to know whether the incentive contract she currently uses to motivate her employees is optimal, and if not, how to improve it. What information must she acquire, and how should she use it to make those assessments? The goal of this paper is to shed light on these questions.

To illustrate our approach, let us begin with a simple example. Suppose that the manager uses a linear contract such as the one illustrated in the left panel of Figure 1, and she wants to determine how she should modify the slope, α, to increase her profits. Let $\Pi = (m - \alpha)a$ denote her profit function, where m denotes her marginal profit, and a denotes the agent's expected output, which of course depends on his effort, and therefore, α. Consider the effects of a marginal increase in α: First, it mechanically increases the the agent's pay, thus reducing the manager's profit, and second, it affects her profit indirectly by inducing the employee to provide a different level of effort. If we differentiate Π with respect to α, we obtain

$$\frac{d\Pi}{d\alpha} = -a + (m - \alpha)\frac{da}{d\alpha} \geq 0 \quad \text{if and only if} \quad \frac{\alpha}{a}\frac{da}{d\alpha} \geq \frac{\alpha}{m - \alpha}\,. \tag{1}$$

Naturally, for the current piece rate to be (locally) optimal, the two effects must exactly offset each other, so that the above inequality binds. If, instead, the inequality is strict, then the manager should increase the slope, whereas she should decrease it if the inequality is violated. Thus, to determine in which direction to perturb the slope of the contract, α, it is necessary and sufficient that the manager knows how the employee's productivity responds to a marginal change in α. The main goal of this paper is to extend this exercise to arbitrary, not necessarily linear contracts.

We consider a canonical principal-agent framework under moral hazard, in the spirit of Holmström (1979). The principal offers a contract to the agent, who then optimally chooses how much effort to provide. We assume that the principal has output data corresponding to some *status quo* contract, as well as to some arbitrary perturbation of this contract, which induces the agent to choose a different effort level and allows the principal to estimate how the agent's productivity responds to a change in his incentives. Her objective is two-fold. First,

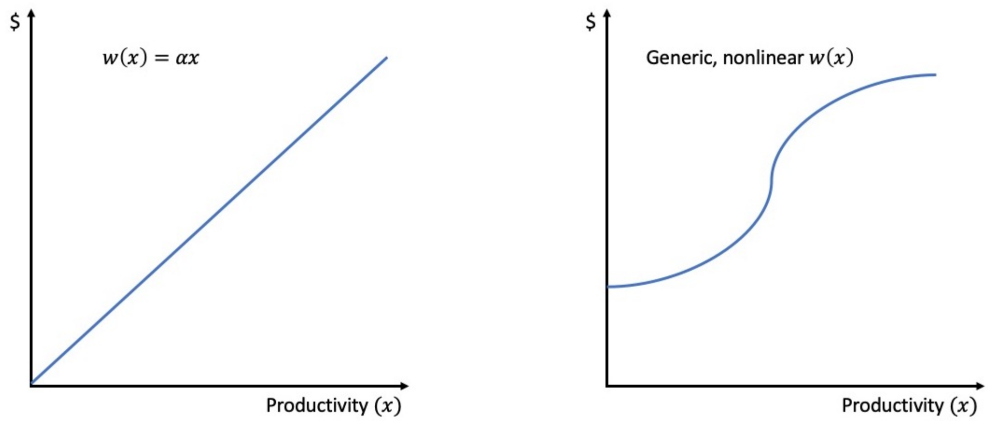

Figure 1: Two Examples

she would like to determine how to optimally perturb the status quo contract to increase her profits; if no profitable perturbation exists, then she can conclude that the status quo contract is locally optimal. The second objective is to find the optimal contract. In doing so, the principal restricts attention to contracts that give the agent at least as much expected utility as the status quo contract, thereby assuming that his participation constraint binds under that contract. A key focus is what information, in a minimal sense, does the principal need to address these questions.

In the above example, the contract consists of one parameter, α, and so knowing how the agent's productivity responds to a single perturbation is sufficient to determine how to optimally perturb the slope of the contract. However, if the principal does not restrict attention to a particular parametric class of contracts, then to determine how to optimally perturb a given contract, say the one illustrated in the right panel of Figure 1, she must predict the agent's response to every possible perturbation, of which there are infinitely many.

Our main lemma shows that if the principal takes a stance on the agent's marginal utility function, then she can predict how the agent would respond to *any* perturbation. To see why this is possible, recall that by assumption, the principal has output data corresponding to a status quo contract and to a perturbation of this contract. Thus, she can estimate the distribution of output corresponding to two effort levels, as well as how this distribution changes in response to a marginal change in effort. In turn, together with an assumption about the agent's marginal utility function, these estimates allow her to predict how any perturbation will affect the agent's marginal incentives. Finally, knowledge of how the agent's effort responds to one change in his marginal incentives is informative about the curvature of his effort cost function, which in turn enables the principal to predict how he would respond to any other perturbation.

Section 4.1 focuses on finding the *best* perturbation. First, let us define *best*. Consider how the principal's expected profit changes as the status quo contract is perturbed in the direction of another, arbitrary, contract. The rate at which her profit changes can be formalized using the Gateaux derivative. We assume that the principal's goal is to find a perturbation that maximizes this derivative, subject to the constraint that it gives the agent at least as much utility as the status quo contract. Intuitively, the principal seeks to perturb the status quo contract in the direction that increases her profits at the fastest rate. This exercise is at the core of various numerical optimization methods, such as the steepest ascent and the gradient ascent algorithms. We show that this problem can be formulated as a standard convex optimization program, and our main result gives an explicit characterization for the optimal perturbation. The status quo contract is (locally) optimal if no profit-increasing perturbation exists.

This result is useful in at least three ways. First, it provides a concrete answer to the question regarding how to modify the status quo contract to increase profits. Second, it sheds light on the informational requirements for making that determination. Specifically, the principal must know (A) the distribution of output corresponding to a status quo contract and a perturbation that induces the agent to (marginally) alter his effort, and (B) the agent's marginal utility function. An implication of (A) is that the principal needs to experiment with the contracts that she uses to motivate her employees. Finally, even if the principal does not have the knowledge specified in (A) or is not willing to take a stance on (B), she can use our formulas to infer what parameters rationalize the status quo contract being optimal.

We then turn to finding the (globally) optimal contract. Our focus is on the informational requirements to determine the optimal contract, rather than the characterization itself, which is identical to Holmström (1979). To do so, the local information described above is insufficient. Instead, the principal must also take a stance on the agent's effort cost function and the distribution of output corresponding to every effort level in the relevant range.

In Section 5, we show that our results generalize to four extensions. The first is to incorporate limited liability constraints. Second, we consider the case in which the principal offers a common contract to agents with heterogeneous abilities. Next, we consider the possibility that the agent's effort is multidimensional; for example, the agent might exert effort towards selling different products, or he might exert effort to increase output, as well as improve quality. To find the optimal perturbation in this case, the principal must know how the agent's N-dimensional effort responds to $\lceil (N+1)/2 \rceil$ different perturbations. Finally, we characterize the optimal perturbation when the principal restricts attention to a particular parametric class of contracts, such as linear or piece-wise linear ones. Perhaps surprisingly, the principal needs no less information than the case in which she does not restrict attention to a particular class.

Finally, Section 6 illustrates how our methodology can be applied using a dataset from the real-effort experiment in DellaVigna and Pope (2017). This dataset contains output data from a large number of subjects carrying out a task on Amazon MTurk under different monetary and behavioral incentives. Starting with the premise that the principal has output data corresponding to two "contracts", we characterize the optimal perturbation, and we discuss some of the intricate choices that a principal must make when faced with this problem. Perhaps surprisingly and reassuringly, we find that within the class of utility functions that exhibit constant relative risk aversion, the principal's assumption about the agent's degree of risk aversion has little impact on the optimal perturbation.

This paper straddles the theoretical and the empirical literatures on principal-agent problems under moral hazard. The canonical model (Mirrlees (1976) and Holmström (1979)) considers a principal who wants to motivate an agent to choose a particular unobservable action; *e.g.,* work hard. To do so, she offers a contract to an agent, which specifies a schedule of payments conditional on the realization of a signal that is correlated with the agent's action. Extensions of this model include settings in which the signal is not contractible, the agent's action is multidimensional and some tasks are easier to measure than others, or the principal and the agent interact repeatedly—see Bolton and Dewatripont (2005) for a comprehensive treatment. The goal of the theoretical literature, typically, is to characterize an optimal contract under the premise that the principal has perfect knowledge of all relevant parameters of the model.

The empirical literature can be classified into (at least) two groups. The first examines the degree to which workers respond to incentives as predicted by the theory. For example, Lazear (2000) finds that the switch from hourly wages to piece-rate pay at Safelite Auto Glass led to a 44% increase in productivity, approximately half of which is attributable to workers exerting more effort, while the other half is due to selection, that is, more productive workers joining the firm and less productive ones leaving. In similar vein, Shearer (2004) finds a 20% increase in productivity when tree planters in British Columbia were paid according to piece rates, compared to fixed wages. See also Paarsch and Shearer (1999) for a related study.[1] Others study work on more complex tasks that are amenable to the multitasking problem;

[1]Oettinger (2001) finds a positive effect of commissions on sales for stadium vendors, and Fehr and Goette (2007) finds a positive effect of commissions on productivity for bicycle messengers in Zürich. Bandiera et al. (2007) and Bandiera et al. (2009) measure the effect of introducing performance pay for managers on their subordinates' productivity. Guiteras and Jack (2018) studies the incentive effect on productivity and selection for labor workers in rural Malawi. Hill (2019) estimates the effect of an increase in the minimum wage on productivity for strawberry pickers in California.

see, for example, Holmström and Milgrom (1991). For example, Gibbs et al. (2017) exploits a field experiment at an Indian technology firm to estimate the impact of introducing explicit incentives for submitting ideas for process improvements. They find that incentives led employees to submit fewer fewer but higher-quality ideas.[2] On a broader scale, Prendergast (2014) uses estimates for the elasticity of income to marginal tax rates (see, for example, Brewer, Saez and Shephard (2010)) to establish an upper bound for the responsiveness of worker productivity to incentives. The second category investigates the extent to which observed contracts are consistent with theoretical models. See Chiappori and Salanié (2003) and the references within.

A limitation of the theoretical literature is that it often assumes *too much* knowledge on the principal's behalf (*e.g.*, about the agent's preferences, the actions at his disposal and the associated cost, and how these actions map into the contractible signal). On the other hand, the empirical literature usually focuses on estimating how different incentive vehicles affect performance. The goal of this paper is to bridge these literatures by exploring how an organization can combine lessons from the theoretical agency literature together with estimates such as those described above to improve its incentive systems.

Our work is conceptually related to papers that use a variational approach to characterize optimal mechanisms in terms of the relevant elasticities. For example, the Lerner index relates the optimal monopoly price to the price elasticity of demand (see, for example, Tirole (1988)), and Wilson (1993) characterizes an optimal quantity-discount price-menu. Saez (2001) and a growing literature derives optimal income tax formulas using elasticities of earnings with respect to tax rates. To the best of our knowledge, we are the first to extend this approach to a principal-agent problem under moral hazard. Methodologically, the key difference is the following: The extant literature has focused on models in which all uncertainty is resolved before the agent chooses his action, and so a local perturbation of the mechanism has a local behavioral effect. In contrast, because uncertainty is resolved after the agent chooses his action in our setting, any perturbation of the contract will generally have a global effect on the agent's incentives.

2 Model

Environment.— We consider the contractual relationship between a principal and one or more homogeneous agents. The principal offers an output-contingent contract $w(x)$ to each

[2]Similarly, Balbuzanov et al. (2017) finds that the introduction of incentives led journalists in Kenya to submit fewer, higher quality articles. Hong et al. (2018) estimates the impact of piece rates at a Chinese manufacturing firm on the quantity and quality of output.

agent, who, after observing the contract, chooses effort $a \geq 0$, which is not contractible. His output, x, is realized according to some cumulative distribution function, $F(x|a)$, with probability density function (hereafter pdf) $f(x|a)$, which we assume are twice differentiable in a. Finally, payoffs are realized and the game ends. Without loss of generality, we normalize a such that $a = \mathbb{E}[x|a]$, so that the agent's effort can be interpreted as his expected output.

Actions.— The principal chooses a contract $w : \mathbb{R} \to \mathbb{R}$, which is an upper-semicontinuous (hereafter, u.s.c) mapping from output to transfers made to the agent. We assume that the principal restricts attention to contracts that leave each agent with at least as much expected utility as some generic, status quo contract, $\widehat{w}$. After observing the contract, each agent chooses an effort $a \geq 0$.

Information.— Each agent knows all parameters that are pertinent to his decisions, that is, he knows his utility function $v(\omega)$, his cost function $c(a)$, and the pdf $f(\cdot|a)$ for every feasible effort level. The principal knows her marginal profit $m > 0$, the distribution of output corresponding to a status quo contract $\widehat{w}$, as well as the distribution corresponding to some perturbed contract, $\widehat{w} + \widehat{t}$, where $\widehat{t}$ is an u.s.c mapping from $\mathbb{R}$ to $\mathbb{R}$. Letting $a(w)$ denote the effort induced by contract w, we assume that $a(\widehat{w} + \widehat{t})$ is sufficiently close but distinct from $a(\widehat{w})$, thus allowing the principal to estimate

$$f_a(x|a(\widehat{w})) := \left.\frac{df(x|a)}{da}\right|_{a=a(\widehat{w})} \simeq \frac{f(x|a(\widehat{w}+\widehat{t})) - f(x|a(\widehat{w}))}{a(\widehat{w}+\widehat{t}) - a(\widehat{w})}\,. \tag{2}$$

For now, we abstain from specifying the principal's knowledge about other parameters.

Preferences.— If an agent is paid ω and exerts effort a, then he obtains utility $v(\omega) - c(a)$, where $v : \mathbb{R} \to \mathbb{R}$ and $c : \mathbb{R}_+ \to \mathbb{R}_+$ are twice continuously differentiable, and satisfy $v'' < 0 < v'$ and $c', c'' > 0$. The agent chooses his effort to maximize his expected utility. If an agent generates output x and is paid $w(x)$, then the principal's corresponding profit is $mx - w(x)$. The principal's objectives are two-fold. The first, and the one we will focus on, is to perturb the status quo contract, $\widehat{w}$, in the direction that increases her expected profit at the fastest rate.[3] Naturally, $\widehat{w}$ is locally optimal if and only if there exists no profit-increasing perturbation. The second goal is to find the profit-maximizing contract.

2.1 Discussion of the Model

Let us discuss some of our assumptions. First, we implicitly assume that the principal uses a performance measure that is sufficiently broad so as to avoid the multitasking problem

[3] We implicitly assume that the principal uses the steepest ascent method to improve the status quo contract; see, for example, Section 9.4 in Boyd and Vandenberghe (2004).

(Baker, 2000). Second, the restriction to contracts that give the agent at least as much utility as the status quo contract is equivalent to assuming that the agent's participation constraint binds under $\widehat{w}$.[4] Third, we assume that the output data corresponding to the perturbed contract $\widehat{w} + \widehat{t}$ is not tainted by ratchet effects—a common concern when firms experiment with their performance pay plans.

Next, we discuss some assumptions that we relax in later sections. In Section 5.1, we incorporate limited liability constraints on behalf of the parties, and in Section 5.2 we extend our analysis to the case in which the agents have heterogeneous abilities. In Section 5.3, we consider the case in which each agent's action is multi-dimensional, and in Section 5.4, we assume that the principal restricts attention to contracts from a specific parametric class; *e.g.*, linear or piece-wise linear contracts. In Appendix A.1, we give conditions such that our results hold (verbatim) if the agent is also motivated by other factors, such as the prospect of a promotion, the threat of firing, prestige, and so on. Finally, in Appendix A.2, we extend our analysis to the case in which the agent's utility is multiplicatively separable in income and effort.[5]

3 Preliminaries

In Sections 3.1 and 3.2, we introduce notation, and we formulate the agent's and the principal's problem, respectively. In Section 3.3, we determine what information the principal needs to know, or otherwise take a stance on, to achieve her objective.

3.1 Agent's Problem

Given contract w, the agent chooses his effort to maximize his expected utility, that is, he chooses

$$a(w) \in \arg\max_{\tilde{a} \geq 0} \left\{ \int u(w(x)) f(x|\tilde{a}) dx - c(\tilde{a}) \right\} .$$

We assume that the first-order approach is valid, so that $a(w)$ satisfies the first-order condition

$$\int v(w(x)) f_a(x|a(w)) dx = c'(a(w)) . \tag{IC}$$

[4] Additionally, when firms revise their performance pay plans, workers are often suspicious about the firm's intentions, which can lead to opposition (*e.g.*, in the form of unionization) and attrition; see, for example, Hall et al. (2000). Restricting attention to contracts that make workers at least as well off as the status quo contract may ease those tensions.

[5] We restrict attention to additively and multiplicatively separable utility functions, first, to rule out random contracts possibly being optimal, and second, because the literature on estimating utility functions has focused on these two classes; see, for example, Barseghyan et al. (2018).

Next, we define for any upper semi-continuous (USC) function $t : \mathbb{R} \to \mathbb{R}$, the Gateaux derivative of the agent's effort when the status quo contract, $\widehat{w}$, is perturbed in the direction of t:

$$\mathcal{D}a(\widehat{w}, t) := \frac{da(\widehat{w} + \theta t)}{d\theta}\bigg|_{\theta=0} = \lim_{\theta \to 0} \frac{a(\widehat{w} + \theta t) - a(\widehat{w})}{\theta}\,.$$

This derivative exists, because by assumption, $\widehat{w}$ and t are USC, and $f(\cdot|a)$ and c are twice-differentiable with respect to a. Using (IC), this derivative can be written in terms of primitives as

$$\mathcal{D}a(\widehat{w}, t) = \frac{\int tv'(\widehat{w})\widehat{f}_a dx}{c''(a(\widehat{w})) - \int v(\widehat{w}) f_{aa}(x|a(\widehat{w}))dx}\,, \tag{3}$$

where, for notational simplicity, we will write $\widehat{f} := f(x|a(\widehat{w}))$ and $\widehat{f}_a := f_a(x|a(\widehat{w}))$, and suppress the argument (x) of the functions $\widehat{w}$ and t. An observation that will become helpful is that $\mathcal{D}a(\widehat{w}, t)$ is linear in t.

3.2 Principal's Problem

The principal's expected profit from offering contract w to each agent is equal to

$$\pi(w) := ma(w) - \int w(x) f(x|a(w))dx\,, \tag{4}$$

where $a(w)$ is given in (IC). For any USC function t, we define the Gateaux derivative of the principal's profit when the status quo contract is perturbed in the direction of t:

$$\mathcal{D}\pi(\widehat{w}, t) := \frac{d\pi(\widehat{w} + \theta t)}{d\theta}\bigg|_{\theta=0} = \lim_{\theta \to 0} \frac{\pi(\widehat{w} + \theta t) - \pi(\widehat{w})}{\theta}\,.$$

This derivative is to be interpreted as the principal's marginal benefit from perturbing the contract $\widehat{w}$ in the direction of (some) perturbation t. It exists for the same reasons that $\mathcal{D}a(\widehat{w}, t)$ does. Using (4), it can be rewritten as

$$\mathcal{D}\pi(\widehat{w}, t) = \left(m - \int \widehat{w}\widehat{f}_a dx\right)\mathcal{D}a(\widehat{w}, t) - \int t\widehat{f} dx\,. \tag{5}$$

This expression has an analogous interpretation as (1): Perturbing the status quo contract has two effects on the principal's profit. First, it induces a change in the agent's effort, as captured by the first term, and second, holding effort fixed, it mechanically affects profits, as captured by the second term. Therefore, a perturbation t is profit-increasing if (and only if) $\mathcal{D}\pi(\widehat{w}, t) > 0$. Finally, we remark that because (3) is linear in t, so is (5).

Recall that the principal restricts attention to perturbations that leave each agent with at least as much expected utility as $\widehat{w}$. Thus, the set of feasible perturbations

$$\mathcal{T} := \left\{ t \, : \, \frac{d}{d\theta} \int v\left(\widehat{w}(x) + \theta t(x)\right) f(x|a(w+\theta t))dx - c(a(w+\theta t)) \bigg|_{\theta=0} \geq 0 \right\}.$$

That is, for any feasible perturbation t, each agent's expected utility must be non-decreasing as the status quo contract is perturbed in the direction of t. Using (IC), the set of feasible perturbations can be rewritten as

$$\mathcal{T} = \left\{ t \, : \, \int t v'(\widehat{w}) \widehat{f} dx \geq 0 \right\}. \tag{6}$$

We will first focus on the principal's first objective, which is to find the u.s.c perturbation that maximizes (5) subject to (6), and return to her second objective in Section 4.2. Given an optimal perturbation, t^*, the principal would replace the status quo contract with

$$w(x) \equiv \widehat{w}(x) + \theta t(x)$$

for an appropriately chosen step size $\theta > 0$.

Because (5) and (6) are linear in t, the principal can maximize her objective simply by setting $t(x) = \infty$ for some x, and $t(x) = -\infty$ for all other x, and hence an optimal perturbation will fail to exist. In such cases, it is standard approach to normalize t by its *length*, by imposing the constraint $\|t\|_p \leq 1$ for some $p \in \{2, 3, ..\}$; see, for example, Section 9.4 in Boyd and Vandenberghe (2004).[6] For now, we take the parameter p as given, and return to its interpretation after Proposition 1. Therefore, the principal's problem can be expressed as the following constrained maximization program:

$$\begin{aligned} \max_{t \text{ u.s.c}} \; & \left(m - \int \widehat{w} \widehat{f}_a dx \right) \mathcal{D}a(\widehat{w}, t) - \int t \widehat{f} dx \\ \text{s.t.} \; & \int t v'(\widehat{w}) \widehat{f} dx \geq 0 \\ & \int |t|^p \, dx \leq 1 \end{aligned} \tag{P}$$

We say that t is an *optimal perturbation* if it solves (P) for some p. Naturally, the status quo contract is (locally) optimal if and only if the optimal perturbation $t(x) \equiv 0$.

[6] It is necessary that $p \geq 2$; if $p = 1$, then the principal would want to make $t(x)$ as large as possible for some $x = \overline{x}$, as small as possible for some $x = \underline{x}$, and set it to zero for all other values of x. In this case, the maximum will not be attained, and additionally, the resulting perturbed contract $\widehat{w} + \theta t$ will differ from $\widehat{w}$ on a set of measure zero and will fail to be u.s.c.

3.3 Sufficient Statistics

In this section, we examine what information the principal must have to solve (P). By assumption, she knows her marginal profit m, the status quo contract $\widehat{w}$, and the corresponding pdf of output, $\widehat{f}$. Moreover, by assumption, she can estimate how an agent's effort and his distribution of output responds to a particular perturbation $\widehat{t}$, that is she can estimate $\widehat{f}_a$ and

$$\mathcal{D}a(\widehat{w},\widehat{t}) = \frac{\int tv'(\widehat{w})\widehat{f}_a dx}{c''(a(\widehat{w})) - \int v(\widehat{w}) f_{aa}(x|a(\widehat{w}))dx} \simeq a(\widehat{w}+\widehat{t}) - a(\widehat{w})\,,$$

where the second term is taken from (3) and the third term follows from a first-order Taylor expansion. In addition to these parameters, to solve (P), the principal must know $\mathcal{D}a(\widehat{w},t)$ for *every* feasible perturbation t, as well as the agent's marginal utility function, v'.

Observe that the perturbation t appears only in the numerator of (3). Therefore, if the principal knows (or takes a stance on) v', then she can use her estimate of $\mathcal{D}a(\widehat{w},\widehat{t})$ to compute the denominator of (3), and consequently, the agent's response $\mathcal{D}a(\widehat{w},t)$ for any other t. This leads us to the following remark.

Remark 1. *For any perturbation t, the Gateaux derivative of the agent's effort when $\widehat{w}$ is perturbed in the direction of t,*

$$\mathcal{D}a(\widehat{w},t) = \frac{\mathcal{D}a(\widehat{w},\widehat{t})}{\int \widehat{t}v'(\widehat{w})\widehat{f}_a dx} \int tv'(\widehat{w})\widehat{f}_a dx\,. \tag{7}$$

Therefore, if principal knows the agent's marginal utility function, v', then she can solve (P).

As can be seen from (3), to know how the agent will respond to an arbitrary perturbation, one must have (local) knowledge of the curvature of his effort cost function, $c''(a)$, and the first two derivatives of the density of output, $\widehat{f}_a$ and $\widehat{f}_{aa}$. Perhaps surprisingly, (7) shows that $\mathcal{D}a(\widehat{w},\widehat{t})$ contains all the information that is necessary, together with v' and $\widehat{f}_a$, to compute $\mathcal{D}a(\widehat{w},t)$ for any other t, and in this sense, they are sufficient statistics for (P).

Intuitively, this is because any two perturbations that induce a marginal change in the agent's effort, will lead to the same change in the distribution of output.[7] So if the principal takes a stance on the agent's marginal utility function, then she can predict how his marginal incentives will change following any perturbation. Finally, the principal can use her information about the agent's response to some perturbation $\widehat{t}$, together with the fact that $\mathcal{D}a(\widehat{w},t)$ is linear in t to infer how he will respond to any other perturbation.

[7] This implies that the distribution of output depends on the agent's effort and not on the offered contract. This observation is no longer true if, for example, agents can manipulate the distribution of output while holding its expectation fixed as in Barron, Georgiadis, and Swinkels (2019).

Next, we use (7) to rewrite the principal's problem, (P), in terms of parameters that can be estimated using observational data, and the agent's marginal utility function, v', which we henceforth assume that the principal knows. Let us define the constant

$$\mu := \frac{\left(m - \int \widehat{w}\widehat{f}_a dx\right) \mathcal{D}a(\widehat{w}, \widehat{t})}{\int \widehat{t}v'(\widehat{w})\widehat{f}_a dx}. \tag{8}$$

Using Remark 1, we can rewrite the principal's problem as

$$\begin{aligned} t^* = \arg \max_{t \text{ u.s.c}} \; & \mu \int tv'(\widehat{w})\widehat{f}_a dx - \int t\widehat{f} dx \\ \text{s.t.} \; & \int tv'(\widehat{w})\widehat{f} dx \geq 0 \\ & \int |t|^p \, dx \leq 1 \end{aligned} \tag{P'}$$

We conclude this section with two additional remarks. First, if the firm has estimated $\mathcal{D}a(\widehat{w}, t)$ for two or more perturbations, then it can use (7) to evaluate whether a particular marginal utility function is consistent with the estimated $\mathcal{D}a$'s. Alternatively, the principal might assume that v belongs to a particular parametric family (*e.g.*, utility functions that exhibit constant relative risk aversion), and use (7) together with the estimated $\mathcal{D}a$'s to infer the unknown parameter(s) of v.

Second, observe that the denominator of (3) is the negative of the second derivative of the agent's expected utility with respect to a. Therefore, one can inspect whether the first-order approach is locally valid at $a(\widehat{w})$, which is necessary for the validity of this analysis, by verifying that $\mathcal{D}a(\widehat{w}, \widehat{t})$ and $\int \widehat{t}v'(\widehat{w})\widehat{f}_a dx$ have the same sign.

4 Main Results

In Section 4.1, we characterize the solution to (P'), and in doing so, we establish a necessary and sufficient condition such that the status quo contract is locally optimal. Then in Section 4.2, we examine the informational requirements for characterizing the optimal contract.

4.1 Optimal Perturbation and Local Optimality

To solve (P') we use the method of Lagrange multipliers. Let $\lambda \geq 0$ and $\nu \geq 0$ denote the dual multipliers associated with the first and second constraint, respectively. Then the

Lagrangian can be written as

$$L(\lambda,\nu)=\max_{t}\left\{\nu+\int\left[t\left(\lambda v'(\widehat{w})\widehat{f}+\mu v'(\widehat{w})\widehat{f}_a-\widehat{f}\right)-\nu\,|t|^p\right]dx\right\}. \tag{9}$$

For any $\nu>0$, we can optimize the integrand with respect to t pointwise. Noting that the integrand is differentiable with respect to t except at $t=0$, the corresponding first-order condition implies that

$$t_{\lambda,\nu}=\operatorname{sgn}\left\{\left(\lambda\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})-\widehat{f}\right\}\left|\frac{\left(\lambda\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})-\widehat{f}}{\nu p}\right|^{\frac{1}{p-1}},$$

where $\operatorname{sgn}\{\cdot\}$ is the sign function, and recall that t, $\widehat{f}$, $\widehat{f}_a$, and $\widehat{w}$ are functions of x.[8] To ensure that $t_{\lambda,\nu}$ is u.s.c, without loss of generality, we adopt the convention that $\operatorname{sgn}\{0\}=1$.

Next, we pin down the optimal multipliers λ and ν, by turning to the dual problem, and solving the following minimization program:

$$\min_{\lambda\geq 0\,,\,\nu\geq 0}L(\lambda,\nu)\,.$$

This problem is convex, and using $t_{\lambda,\nu}$, the corresponding first-order conditions yield

$$\lambda^*=\min\left\{\lambda'\geq 0:\int sgn\left\{\left(\lambda'\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})-f\right\}\left|\left(\lambda'\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})-f\right|^{\frac{1}{p-1}}v'(\widehat{w})\widehat{f}\,dx\geq 0\right\} \tag{10}$$

and

$$\nu^*=\frac{1}{p}\left[\int\left|\left(\lambda\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})-\widehat{f}\right|^{\frac{p}{p-1}}dx\right]^{\frac{p-1}{p}}. \tag{11}$$

The following proposition proves that the perturbation t_{λ^*,ν^*} is indeed optimal (*i.e.,* strong duality holds), and gives a necessary and sufficient condition such that the status quo contract is locally optimal.

Proposition 1. *The status quo contract, $\widehat{w}$, is locally optimal if and only if*

$$\lambda+\mu\frac{\widehat{f}_a}{\widehat{f}}=\frac{1}{v'(\widehat{w})} \tag{12}$$

for all x, where $\lambda=\int\widehat{f}/v'(\widehat{w})dx$ and μ is given in (8).

[8]If $\nu=0$, then the integrand of (9) is linear in t, and the first-order condition implies that $\left(\lambda\widehat{f}+\mu\widehat{f}_a\right)v'(\widehat{w})=\widehat{f}$, and hence $L(\lambda,0)=0$.

Otherwise, the optimal perturbation

$$t^* = \frac{\left[v'(\widehat{w})\widehat{f}\right]^{\frac{1}{p-1}}}{\left[\int \left|\left(\lambda\widehat{f} + \mu\widehat{f}_a\right) v'(\widehat{w}) - \widehat{f}\right|^{\frac{p}{p-1}} dx\right]^{\frac{1}{p}}} sgn\left(\lambda + \mu\frac{\widehat{f}_a}{\widehat{f}} - \frac{1}{v'(\widehat{w})}\right)\left|\lambda + \mu\frac{\widehat{f}_a}{\widehat{f}} - \frac{1}{v'(\widehat{w})}\right|^{\frac{1}{p-1}} \tag{13}$$

where μ and λ are given in (8) and (10), respectively.

The optimality condition in (12) is familiar from the canonical principal-agent model under moral hazard; see, for example, Holmström (1979). We remark, however, that the dual multipliers λ and μ given in Proposition 1 *contain information* that is not contained in the standard solutions. To be specific, it is standard procedure to solve the canonical principal-agent model in two stages; see, for example, Grossman and Hart (1983). First, one fixes an arbitrary *target* effort level, and finds the contract that maximizes the principal's objective subject to the agent's participation and incentive compatibility constraint (*i.e.,* the constraint that the agent finds it optimal to choose the target effort level). It is well-known that for any fixed effort level, the profit-maximizing contract satisfies (12) for *some* dual multipliers λ and μ. These multipliers are chosen such that there exists no perturbation that increases the principal's profit, while holding the agent's utility and his optimal effort choice constant. In contrast, the multipliers characterized in Proposition 1 also consider perturbations in which the agent changes his effort level.

Next, let us turn to the optimal perturbation. Observe from (13) that t^* specifies that the principal should raise payments to the agent at every x such that $\left(\lambda\widehat{f} + \mu\widehat{f}_a\right) v'(\widehat{w}) > \widehat{f}$, and decrease payments if the reverse inequality holds. Marginally increasing payments at x has three effects on the principal's profit: First, it relaxes the constraint that the perturbation must give the agent at least as much utility as $\widehat{w}$, which has implicit value $\lambda v'(\widehat{w})\widehat{f}$. Second, it affects the agent's effort which has implicit value $\mu v'(\widehat{w})\widehat{f}_a$. Finally, holding effort constant, it reduces the principal's profit at rate $\widehat{f}$. Thus, the optimal perturbation increases payments at x if the principal's marginal benefit of raising payments at x outweighs the respective marginal cost. The amount by which payments are changed depends on the choice of the parameter p.

This parameter changes how the optimal perturbation reacts to the principal's marginal value of increasing payments at a given level of output, which is proportional to $\lambda + \mu\widehat{f}_a/\widehat{f} - 1/v'(\widehat{w})$. At the one extreme, if $p = 2$, then t^* is linear in this quantity. On the other hand, as $p \to \infty$, $t^* = 1$ if this quantity is positive, and $t^* = -1$ if it is negative. Therefore, if the principal wants to make the perturbation very sensitive to the informativeness of $\widehat{f}_a/\widehat{f}$, then she should set p close to two, whereas if she wants to make the perturbations sensitive only

m	Principal's marginal profit
$\mathcal{D}a(\widehat{w},\widehat{t})$	Marginal change in effort when $\widehat{w}$ is perturbed in the direction of $\widehat{w}+\widehat{t}$
$f(\cdot\|a(\widehat{w}))$	Pdf of output when $a=a(\widehat{w})$
$f_a(\cdot\|a(\widehat{w}))$	Marginal change in pdf of output due to change in effort at $a=a(\widehat{w})$
$v'(\cdot)$	Agent's marginal utility for money

Table 1: Sufficient statistics to find optimal perturbation of the status quo contract, $\widehat{w}$.

to whether $\widehat{f}_a/\widehat{f}$ is bigger than some cutoff, then she should set a large p.

Proposition 1 is valuable in three ways. First, it informs us which parameters the firm must take a stance on to determine how to optimally perturb the status quo contract. Table 1 lists these sufficient statistics. Second, it provides an explicit formula for identifying the perturbation that increases the principal's profit at the fastest rate. Specifically, the principal should replace the status quo contract, $\widehat{w}$, with

$$\widehat{w}^* := \widehat{w} + \theta t^*,$$

for some appropriately judged step size $\theta > 0$. For an overview of different ways to determine the appropriate step size, see Boyd and Vandenberghe (2004).

Finally, (12) can be used to infer what assumptions rationalize $\widehat{w}$ being optimal. For instance, consider a firm who is unwilling to experiment with its incentive contract, and thus cannot estimate $\mathcal{D}a(\widehat{w},t)$ for any t.[9] Nevertheless, such a firm can use (12) to infer what effort responses, $\mathcal{D}a(\widehat{w},t)$, are consistent with $\widehat{w}$ being optimal, and evaluate the extent to which they are reasonable. Alternatively, the principal might assume that the agent's utility function belongs to a parametric family, and under the premise that $\widehat{w}$ is optimal, estimate the unknown parameters.

We remark two caveats related to (12). First, even if $\widehat{w}$ satisfies (12) for all x, it need not be globally optimal. For a locally optimal contract to also be globally optimal, one has to make strong assumptions on the parameters; see Section 4 in Jewitt, Kadan, and Swinkels (2008) for details. Second, (12) can be used to verify whether a particular contract is optimal, but it cannot be used to characterize the optimal contract, for example, by finding a contract w which satisfies (12) for all x. The reason is that the multiplier μ depends the specific $\widehat{w}$ and $\widehat{t}$. For a characterization of the optimal contract, see Section 4.2.

[9] For example, Lincoln Electric is infamous for not experimenting with its performance pay plans in fear of ratchet effects (Hall et al., 2000).

4.2 Optimal Contract

In this section, we turn to the characterization of the (globally) optimal contract. This is a canonical principal-agent model, as analyzed by Holmström (1979) and others. Thus, our focus is not on the characterization itself, but rather, on the parameters that the principal must take a stance on to find the optimal contract.

As discussed in the previous subsection, it is standard approach to solve this problem in two stages (see, for example, Grossman and Hart (1983)). The first stage entails characterizing, for every feasible effort level, the cost-minimizing contract subject to the agent's participation and incentive compatibility constraint. In our setting, by assumption, any contract must give the agent at least as much utility as $\widehat{w}$. In the second stage, the principal solves for the profit-maximizing effort, and the corresponding contract is (globally) optimal.

Let us begin by considering the first-stage problem. To do so, fix an (arbitrary) effort level a. The constraints that the contract must give the agent at least as much utility as $\widehat{w}$ and induce him to exert effort a can be written as

$$\int v(w(x))f(x|a)dx \geq \int v(\widehat{w})\widehat{f}dx + \int_{a(\widehat{w})}^{a} c'(s)ds \text{ , and} \tag{14}$$

$$\int v(w(x))f_a(x|a)dx \geq c'(a)\,, \tag{15}$$

respectively. Thus, for fixed effort a, the principal solves the following constrained optimization program:

$$\min_{w}\left\{\int w(x)f(x|a)dx \,:\, (14) \text{ and } (15)\right\}. \tag{16}$$

The following proposition, due to Jewitt, Kadan, and Swinkels (2008), characterizes the solution to program.

Proposition 2 (Jewitt, Kadan, and Swinkels (2008))**.** *Suppose that a solution to (16) exists, denoted by w_a. Then there exist constants $\lambda_a \geq 0$ and $\mu_a \geq 0$ which satisfy*

$$\frac{1}{v'(w_a(x))} = \lambda_a + \mu_a\frac{f_a(x|a)}{f(x|a)} \tag{17}$$

for every x, and the respective complementary slackness conditions; i.e., *either $\lambda_a = 0$ (resp. $\mu_a = 0$) or (14) (resp. (15)) binds.*

m	Principal's marginal profit
$f(\cdot\|a)$	Pdf of output corresponding to every feasible a
$v'(\cdot)$	Agent's marginal utility for money
$c'(a)$	Agent's marginal cost of effort for every feasible a

Table 2: Sufficient statistics for characterizing the optimal contract.

To find the profit-maximizing effort, the principal then solves

$$a^* \in \arg\max_a \left\{ ma - \int w_a(x) f(x|a) dx \right\} , \tag{18}$$

and provided that the first-order approach is valid, w_{a^*} is the globally optimal contract (among those that give the agent at least as much utility as $\widehat{w}$).

To find the optimal contract, the principal must solve (17) for every feasible effort level a, which requires that for every effort level a (in the relevant range), she takes a stance on the pdf of output and its derivative, $f(\cdot|a)$ and $f_a(\cdot|a)$, as well the agent's marginal cost of effort $c'(a)$. Table 2 provides a summary. We conclude this section with two remarks. First, if the principal has only *local* information at effort level a, such as that listed in Table 1, then she can still characterize the cost-minimizing contract w_a. Second, the principal can use her local knowledge of $\widehat{f}$, $\widehat{f}_a$, and $\mathcal{D}(\widehat{w}, \widehat{t})$ to ensure that her global assumptions about f and c' are internally consistent.

5 Extensions

In this section, we consider four extensions of the baseline model. First, we incorporate limited liability constraints. Second, we allow the agents' effort costs to be heterogeneous. Third, we consider the possibility that each agent's effort is multidimensional, and finally, we consider the case in which the principal restricts attention to a particular parametric class of contracts.

5.1 Bounded Payments

We consider the principal's problem of finding an optimal perturbation when the payments she can make to the agent are bounded, for example by limited liability and liquidity constraints. Specifically, we assume that any contract has to specify payments to the agents that are no smaller than some minimum wage $\underline{w}$, and no larger than some upper bound, $\overline{w}$. Moreover, assume that this constraint binds for some levels of output under the status quo contract. Given a fixed step size $\theta > 0$, the principal's problem can be expressed as

$$\begin{aligned}
\max_{t \text{ u.s.c}} \ & \mu \int t v'(\widehat{w}) \widehat{f}_a dx - \int t \widehat{f} dx \\
\text{s.t.} \ & \int t v'(\widehat{w}) \widehat{f} dx \geq 0 \\
& \int |t|^p \, dx \leq 1 \\
& \widehat{w}(x) + \theta t(x) \geq \underline{w} \quad \forall x \\
& \widehat{w}(x) + \theta t(x) \leq \overline{w} \quad \forall x
\end{aligned}$$

where μ is given in (8) and $p \in \{2, 3, ..\}$. That is, the perturbed contract $\widehat{w} + \theta t$ must satisfy the liability constraints for every level of output. Since the additional constraints relative to (P') are linear, the problem remains a convex optimization program, and it can be solved using the same approach as (P').

5.2 Heterogeneous Abilities

So far, we have assumed that the principal contracts with one or more homogeneous workers. Real firms, of course, employ heterogeneous workers, and one interpretation of the exercise is that the firm offers different contracts to workers of different types. In practice, however, firms typically offer a single contract to all its workers in a given job. The goal of this section is to extend the methodology developed in Sections 3.1-4.1 when the agents have heterogeneous effort costs.

We consider the same model as in Section 2, except that each agent is indexed by a type $\phi \in \mathbb{N}$. Agents with different types have different cost of efforts but are otherwise identical. We denote the cost of a type-ϕ agent choosing effort a (which generates expected output equal to a) by $c(a, \phi)$. Assume that the principal has output data corresponding to some status quo contract, $\widehat{w}$, and a perturbation of this contract, denoted by $\widehat{w} + \widehat{t}$. Also assume that this data allows the principal to classify the agents into types, and for each type ϕ, estimate the density of output and its derivative, $\widehat{f}^\phi := f(\cdot|a^\phi(\widehat{w}))$ and $\widehat{f}_a^\phi := f_a(\cdot|a^\phi(\widehat{w}))$, where $a^\phi(w)$ denotes the effort chosen by a type-ϕ agent when offered contract w, as well as

$$\mathcal{D}a^\phi(\widehat{w}, \widehat{t}) := \left. \frac{da^\phi(\widehat{w} + \theta \widehat{t})}{d\theta} \right|_{\theta=0} = \lim_{\theta \to 0} \frac{a^\phi(\widehat{w} + \theta \widehat{t}) - a^\phi(\widehat{w})}{\theta}.$$

The principal's goal is to perturb the status quo contract in the direction that increases her profit at the fastest rate, subject to the constraint that it gives each type at least as much utility as $\widehat{w}$.

It is straightforward to replicate the analysis in Sections 3.2-3.3 verbatim, with the exception that the expressions in (6) and (7) are indexed by the type, ϕ, and the expected principal's profit takes the average across different types. As a result, the principal's problem can be written as

$$\begin{aligned} \max_{t \text{ u.s.c}} \ & \sum_{\phi} p^{\phi} \left[\mu^{\phi} \int t v'(\widehat{w}) \widehat{f}_a^{\phi} dx - \int t \widehat{f}^{\phi} dx \right] \\ \text{s.t.} \ & \int t v'(\widehat{w}) \widehat{f}^{\phi} dx \geq 0 \ \ \forall \phi \\ & \int |t|^p \, dx \leq 1 \end{aligned} \tag{19}$$

where $p \geq 2$, and for each ϕ, the constant $\mu^{\phi} := \left[m - \int \widehat{w} \widehat{f}_a^{\phi} dx \right] \mathcal{D}a^{\phi}(\widehat{w}, \widehat{t}) \ / \int \widehat{t} v'(\widehat{w}) \widehat{f}_a^{\phi} dx$. Notice that this is a convex maximization problem, and hence it can be solved using the same approach as (P').

It is well known that the design of performance pay may be used to induce *selection*, that is, attract more productive workers and induce less productive workers to exit (Lazear, 2000). To do so in our setting, the principal may restrict attention to perturbations that give at least as much utility as $\widehat{w}$ to only a subset of the more productive types. Formally, this would imply that (19) must hold only for the types that the principal wants to attract, and must be violated for the types that she wants to repel from the firm or dissuade from joining. A complete analysis of the selection effects associated with performance pay is beyond the scope of this paper, and is left for future work.

5.3 Multi-dimensional Effort

In this section, we extend our baseline model to the case in which the agent's effort is multi-dimensional. As an example, each agent might be a salesperson, who can exert effort towards selling different products. To be specific, suppose that for each product $i \in \{1, ..., N\}$, he chooses effort a_i, which in turn generates sales $\mathbf{x} \sim f(\cdot|\mathbf{a})$, where $\mathbf{a}$ and $\mathbf{x}$ is a (finite) vector of efforts and sales, respectively. The cost of choosing $\mathbf{a}$ is equal to $c(\mathbf{a})$, and the function c satisfies the usual conditions. A contract w is an u.s.c mapping from sales $\mathbf{x}$ to a monetary transfer to the agent. We continue to assume that the principal's performance measure, $\mathbf{x}$, is sufficiently broad so that the agent cannot distort it (Baker, 2000).

We assume that the principal has output data corresponding to a status quo contract $\widehat{w}$, and thus she can estimate the density $\widehat{f} := f(\cdot|\mathbf{a}(\widehat{w}))$, where $\mathbf{a}(w)$ denotes the (vector of) efforts chosen by an agent when offered contract w. Given any perturbation t, which is an u.s.c mapping from $\mathbf{x}$ to a real number, let us define for each i, the Gateaux derivative

$$\mathcal{D}a_i(\widehat{w},t) := \left.\frac{da_i(\widehat{w}+\theta t)}{d\theta}\right|_{\theta=0} = \lim_{\theta\to 0}\frac{a_i(\widehat{w}+\theta t)-a_i(\widehat{w})}{\theta}\,. \tag{20}$$

The principal's expected profit when she offers contract w,

$$\pi(w) = \sum_i m_i a_i(w) - \int w(\mathbf{x}) f(\mathbf{x}|\mathbf{a}(w))d\mathbf{x}\,,$$

where m_i is the principal's marginal profit associated with a_i. As in Section 2, her objective is to perturb the status quo contract in the direction that increases her profit at the fastest rate; *i.e.*, solve the following maximization program

$$\max_{t \text{ u.s.c}} \mathcal{D}\pi(\widehat{w},t) = \sum_{i=1}^{m}\left(m_i - \int \widehat{w}\widehat{f}_i d\mathbf{x}\right)\mathcal{D}a_i(\widehat{w},t) - \int t\widehat{f}d\mathbf{x} \tag{21}$$

subject to the constraints that t gives the agent at least as much utility as $\widehat{w}$, and $\|t\|_p \le 1$ for some $p \in \{2,3,..\}$. Note that f_i denotes the derivative of f with respect to a_i, and we have dropped the dependence on $\mathbf{x}$ and on the agent's effort for notational simplicity. We will show that to solve this problem, the principal must be able to estimate $\mathcal{D}a_i(\widehat{w},\widehat{t}_k)$ for at least K perturbations, $\widehat{t}_1,..,\widehat{t}_K$, where $K \ge \lceil (N+1)/2 \rceil$.

Similar to Section 3.1, we assume that the first-order approach is valid, and so given contract w, the agent's effort satisfies for each i

$$\int v\left(w(\mathbf{x})\right) f_i(\mathbf{x}|\mathbf{a}(w))\, d\mathbf{x} = c_i(\mathbf{a}(w))\,, \tag{22}$$

where c_i denotes the derivative of c with respect to its i^{th} argument. Using (22), the constraint that any perturbation t must give the agent at least as much utility as the status quo contract can be written as

$$\int\cdots\int tv'(\widehat{w})\widehat{f}\, d\mathbf{x} \ge 0\,.$$

Using (22), we can compute $\mathcal{D}a_i(\widehat{w},t)$ for each i in terms of primitives as

$$\underbrace{\left[c_{ii} - \int v(\widehat{w})\widehat{f}_{ii}d\mathbf{x}\right]}_{=:B_{ii}}\mathcal{D}a_i(\widehat{w},t) + \sum_{j\neq i}\underbrace{\left[c_{ij} - \int v(\widehat{w})\widehat{f}_{ij}d\mathbf{x}\right]}_{=:B_{ij}}\mathcal{D}a_j(\widehat{w},t) = \underbrace{\int tv'(\widehat{w})\widehat{f}_i d\mathbf{x}}_{=:A_i}\,.$$

This is the counterpart of (3) if the agent's effort is N-dimensional. To solve (21), it suffices that the principal can evaluate $\mathcal{D}a_i(\widehat{w},t)$ for every i and any perturbation t. To do so, she must (i) take a stance on the agent's marginal utility function, v', which will allow her to evaluate the vector A for any perturbation t, and (ii) estimate the matrix B. Observe that the B is symmetric, and so it contains $N(N+1)/2$ unknowns. Therefore, to be able to estimate B, the principal must have output data corresponding to at least $K \geq \lceil (N+1)/2 \rceil$ perturbations, which will then enable her to estimate $\mathcal{D}a_i(\widehat{w}, \widehat{t}_k)$ for each i and k. In that case, it is straightforward to verify that (21) can be solved using the same method as (P').

5.4 Parametric Classes of Contracts

Firms often restrict attention to a particular class of contracts. For instance, linear contracts are very common, as are piece-wise linear and single-bonus contracts. Such contracts may be chosen due to their simplicity, or because of considerations outside of our model.[10]

In this section, we address the same questions as in Section 4.1 assuming that the principal only considers contracts which belong to a particular parametric class, denoted by w_α, where $\alpha \in \mathbb{R}^n$ is a vector of parameters and $n \in \mathbb{N}$. For example, if the principal considers only linear contracts, then $w_\alpha(x) = \alpha_1 + \alpha_2 x$, and she chooses the parameters α_1 and α_2. If she restricts attention to piece-wise linear contracts with a guaranteed minimum wage, then $w_\alpha(x) = \alpha_1 + \alpha_2(x - \alpha_3)^+$ for some α_1, α_2, and α_3 that remain to be chosen.

We maintain the assumptions imposed in Section 2, with the exception that both the status quo contract and the perturbed contract for which the principal has output data belong to the particular parametric class, with parameters $\widehat{\alpha}$ and $\widehat{\alpha} + \widehat{z}$, respectively. The principal's objective is to perturb the parameters of the status quo contract in the direction that increases her profit at the fastest rate subject to the constraint that the agent receives no less expected utility than under $w_{\widehat{\alpha}}$.

Towards this goal, given an arbitrary perturbation vector $z \in \mathbb{R}^n$, we will consider contracts of the form $w_{\widehat{\alpha}+\theta z}$. Following the analysis in Sections 3.1 and 3.2, and abusing notation, we define the directional derivatives

$$\mathcal{D}a(w_{\widehat{\alpha}}, z) := \left. \frac{da(w_{\widehat{\alpha}+\theta z})}{d\theta} \right|_{\theta=0} \quad \text{and} \quad \mathcal{D}\pi(w_{\widehat{\alpha}}, z) := \left. \frac{d\pi(w_{\widehat{\alpha}+\theta z})}{d\theta} \right|_{\theta=0},$$

which we assume are well-defined. Using the agent's incentive compatibility condition, (IC),

[10] For example, if gaming is a concern, then linear contracts may be optimal; see for example, Holmström and Milgrom (1987) and Barron, Georgiadis, and Swinkels (2019). If the worker is expectation-loss averse or the performance measure is endogenous, then single-bonus contracts may be optimal; see for example, Herweg et al. (2010) and Georgiadis and Szentes (2019).

the response of the agent's effort to a perturbation in the direction of z can be written as

$$\mathcal{D}a(w_{\widehat{\alpha}}, z) = \frac{\sum_{i=1}^{n} z_i \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f}_a dx}{c''(a) - \int v(w_{\widehat{\alpha}}) f_{aa}(x|a(w_{\widehat{\alpha}})) dx}.$$

Noting that the perturbation vector z appears only in the numerator, it is straightforward to verify that if the principal knows $\mathcal{D}a(w_{\widehat{\alpha}}, \widehat{z})$ for a particular perturbation vector $\widehat{z}$ and the agent's marginal utility function, v', then she can estimate $\mathcal{D}a(w_{\widehat{\alpha}}, z)$ for any other z.

Next, let us turn to the principal's problem. Using (4), we have

$$\mathcal{D}\pi(w_{\widehat{\alpha}}, z) = \sum_{i=1}^{n} z_i \left[\mu_{\widehat{z}} \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f}_a dx - \frac{\partial}{\partial \alpha_i} \int w_{\widehat{\alpha}} \widehat{f} dx \right], \text{ where } \mu_{\widehat{z}} := \frac{\left(m - \int w_{\widehat{\alpha}} \widehat{f}_a dx \right) \mathcal{D}a(w_{\widehat{\alpha}}, \widehat{z})}{\sum_{i=1}^{n} \widehat{z}_i \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f}_a dx}$$

is a parameter that depends on v', the status quo contract, and the perturbation $\widehat{z}$. The constraint that the principal restricts attention to contracts that give the agent at least as much expected utility as the status quo contract implies that any feasible perturbation vector z must satisfy

$$\sum_{i=1}^{n} z_i \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f} dx \geq 0.$$

Thus, the principal solves the following constrained maximization problem:

$$\begin{aligned} \max_{\mathbf{z} \in \mathbb{R}^n} \; & \sum_{i=1}^{n} z_i \left[\mu_{\widehat{z}} \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f}_a dx - \frac{\partial}{\partial \alpha_i} \int w_{\widehat{\alpha}} \widehat{f} dx \right] \\ \text{s.t. } & \sum_{i=1}^{n} z_i \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f} dx \geq 0 \\ & \sum_{i=1}^{n} |z_i|^p \leq 1 \end{aligned} \tag{23}$$

First, note that to solve (23), the principal must have the same knowledge about the parameters of the problem that she needs to solve the more general problem given in (P'), as summarized in Table 1. In other words, perhaps surprisingly, restricting attention to a parametric class of contracts does not reduce the information requirements for finding the optimal perturbation.

Second, observe that unlike (P'), a solution to (23) may exist even if $p = 1$. Nevertheless, in this case, the problem can be easily reformulated as a linear program, and so it is straightforward to solve. The following proposition gives an explicit characterization when $p = 2$. (The solution when $p \geq 3$ is similar.)

Proposition 3. *Consider the maximization problem given in (23) and assume $p = 2$. The optimal perturbation satisfies, for each i,*

$$z_i^* = \frac{1}{2\nu}\left[\lambda\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}dx + \mu_{\widehat{z}}\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}_a dx - \frac{\partial}{\partial\alpha_i}\int w_{\widehat{\alpha}}\widehat{f}dx\right] \tag{24}$$

where

$$\lambda = \left(-\frac{\sum_{i=1}^n\left[\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}dx\right]\left[\frac{\partial}{\partial\alpha_i}\int w_{\widehat{\alpha}}\widehat{f}dx - \mu_{\widehat{z}}\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}_a dx\right]}{\sum_{i=1}^n\left[\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}dx\right]^2}\right)^+$$

and

$$\nu = \frac{1}{2}\sqrt{\sum_{i=1}^n\left[\lambda_{\widehat{z}}\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}dx + \mu_{\widehat{z}}\frac{\partial}{\partial\alpha_i}\int v(w_{\widehat{\alpha}})\widehat{f}_a dx - \frac{\partial}{\partial\alpha_i}\int w_{\widehat{\alpha}}\widehat{f}dx\right]^2}.$$

The status quo contract, $w_{\widehat{\alpha}}$ is (locally) optimal if and only if $\nu = 0$.

This proposition asserts, first, that the status quo contract if locally optimal if (and only if) $z^* \equiv 0$, and second, that the principal should replace the status quo contract with

$$w^* := w_{\widehat{\alpha}+\theta z^*}$$

for an appropriately judged step size $\theta > 0$.

6 Implementation

The goal of this section is to illustrate how one can apply the techniques developed in the previous sections using a dataset from DellaVigna and Pope (2017). They present the findings from a real-effort experiment on Amazon MTurk, in which subjects were tasked with alternating 'a' and 'b' keystrokes during a ten-minute interval as rapidly as they could. The authors examined the effect of different monetary and behavioral incentives on the subjects' performance. Table 3 summarizes seven of the monetary "contracts" that the authors implemented, where x denotes the number of 'a-b' keystrokes (during the ten-minute interval) and N denotes the sample size corresponding to each contract. For example, the third contract, $w(x) = 100 + 0.01x$, specifies that a subject receives \$1 irrespective of his performance, plus 0.01¢ for each 'a-b' keystroke.

For our analysis, we shall assume that the principal's gross profit margin, $m = 0.2$¢ (per 'a-b' keystroke). Given these parameters and the summary statistics in Table 3, it is easy to verify that the third contract generates the largest (average) profit for the principal. We will assume that the principal has output data corresponding to the third and the fourth contract

Contract (in ¢)	Mean # of keystrokes	Std. Errors	N
$w(x) = 100$	1521	31.23	540
$w(x) = 100 + 0.001x$	1883	28.61	538
$w(x) = 100 + 0.01x$	2029	27.47	558
$w(x) = 100 + 0.04x$	2132	26.42	566
$w(x) = 100 + 0.10x$	2175	24.28	538
$w(x) = 100 + 40\,\mathbb{I}_{\{x \geq 2000\}}$	2136	24.66	545
$w(x) = 100 + 80\,\mathbb{I}_{\{x \geq 2000\}}$	2188	22.99	532

Table 3: Monetary incentive treatments in the experiment in DellaVigna and Pope (2017)

in Table 3. We will designate the former as the *status quo* contract, and so we define

$$\widehat{w}(x) := 100 + 0.01x \quad \text{and} \quad \widehat{t}(x) := 0.03x\,.$$

Assume that the principal's goal is to perturb this contract in the direction that increases her profit at the fastest rate.

As this dataset contains a single data point for each subject (*i.e.,* each subject participated in a single treatment, once), we cannot classify the subjects into different types. Therefore, we will treat them as homogeneous. We will assume that if a subject chooses "effort" a, then the number of 'a-b' keystrokes that he accomplishes during the 10-minute interval, x, is drawn from a probability distribution with expected value a.[11] Recall that the principal has output data corresponding to $\widehat{w}$ and $\widehat{w} + \widehat{t}$. Figure 2 illustrates the empirical distributions corresponding to $a(\widehat{w})$ and $a(\widehat{w} + \widehat{t})$.

The first step is to estimate the pdfs $f(\cdot|a(\widehat{w}))$ and $f(\cdot|a(\widehat{w} + \widehat{t}))$. A standard method for doing so is to use kernels (see, for example, Hansen (2009)). Then we can estimate the derivative of $f(\cdot|a(\widehat{w}))$ with respect to a as follows:

$$f_a(x|a(\widehat{w})) = \frac{d}{d\theta} f(x|a(\widehat{w} + \theta\widehat{t}))\bigg|_{\theta=0} \simeq \frac{f(x|a(\widehat{w} + \widehat{t})) - f(x|a(\widehat{w}))}{a(\widehat{w} + \widehat{t}) - a(\widehat{w})}\,.$$

Figure 3 illustrates the estimated pdf's using the triweight kernel (other kernels yield similar curves), as well as the *score*, $f_a(x|a(\widehat{w}))/f(x|a(\widehat{w}))$. The score has a natural interpretation: positive values are indicative that the agent choose a higher effort than $a(\widehat{w})$, and vice versa. Therefore, the green curve in Figure 3 indicates that values of $x \lesssim 1000$ are strongly informative that the subject chose $a < a(\widehat{w})$, values of $x \in (1000, 2500)$ are relatively uninformative,

[11] A natural concern is that subjects may revise their effort over the course of a 10-minute interval depending on their progress. As DellaVigna and Pope (2017) note, there is little evidence for such behavior under any of the linear contracts. Also note that the assumption that the distribution of output following effort a has mean a is without loss of generality.

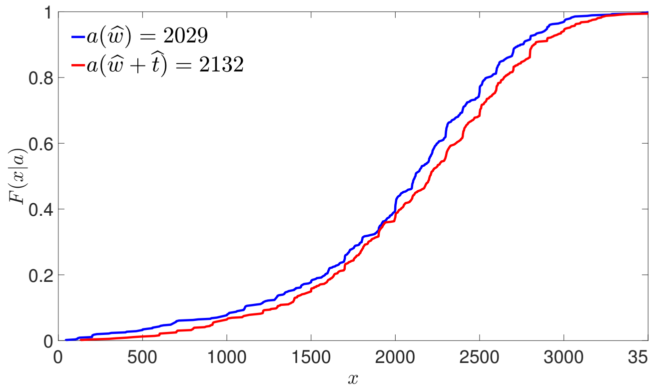

Figure 2: Empirical distribution of x

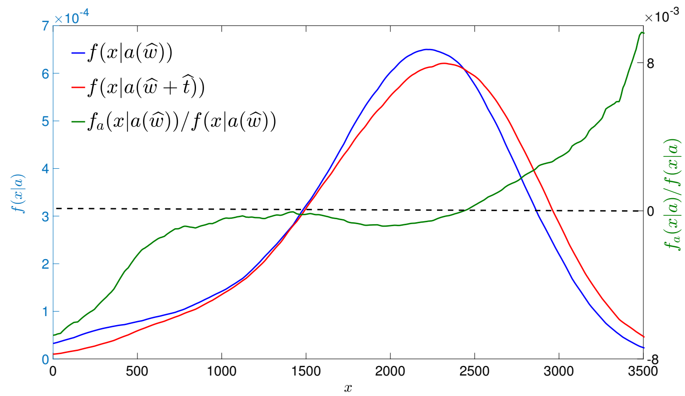

Figure 3: Estimated pdf

and values of $x \gtrsim 2500$ are strongly informative that the subject chose $a > a(\widehat{w})$. Finally, using a first-order Taylor expansion, we can estimate

$$\mathcal{D}a(\widehat{w},\widehat{t}) = \frac{d}{d\theta}a(\widehat{w}+\theta\widehat{t})\bigg|_{\theta=0} \simeq a(\widehat{w}+\widehat{t}) - a(\widehat{w}) = 103\,.$$

Next, we turn to characterizing the optimal perturbation by solving (P'). Recall that to do so, the principal must take a stance on the agent's marginal utility function. We will assume that the agent's utility exhibits constant relative risk aversion with coefficient $\rho \in [0,1]$, and so it is of the form $v'(\omega) = \omega^{-\rho}$. For now, we will set the parameter $p = 2$, and revisit this choice at the end of this section. Figure 4 illustrates the optimal perturbation, t^*, corresponding to four different values of the coefficient of relative risk aversion, ρ. Perhaps surprisingly, ρ has a small impact on the optimal perturbation, which suggests that within this class of utility functions, the principal's assumption about ρ is relatively unimportant.

We now turn to the design of a new contract based on the optimal perturbation. In doing so, it is likely that the principal will be averse to perturbations that lead to a non-monotone contract, or to a contract that is very sensitive to the score, $\widehat{f_a}/\widehat{f}$. Thus, guided by the optimal perturbation given by Proposition 1, she might choose a perturbation such as the one plotted by the dashed line in Figure 4, $\widetilde{t}$.[12] That is, she might choose the new contract

$$\widetilde{w}(x) := \widehat{w}(x) + \theta\widetilde{t}(x)\,,$$

for some $\theta > 0$. The choice of θ should be determined by the principal's confidence with

[12]The "simple" perturbation, $\widetilde{t}$, is projected to increase the principal's profit, as measured by $\mathcal{D}\pi(\widehat{w},t)$, at nearly the same rate as the optimal perturbation, t^*. Specifically, simulations indicate that $\mathcal{D}\pi(\widehat{w},\widetilde{t}) \gtrsim 95\% \times \mathcal{D}\pi(\widehat{w},\widehat{t})$.

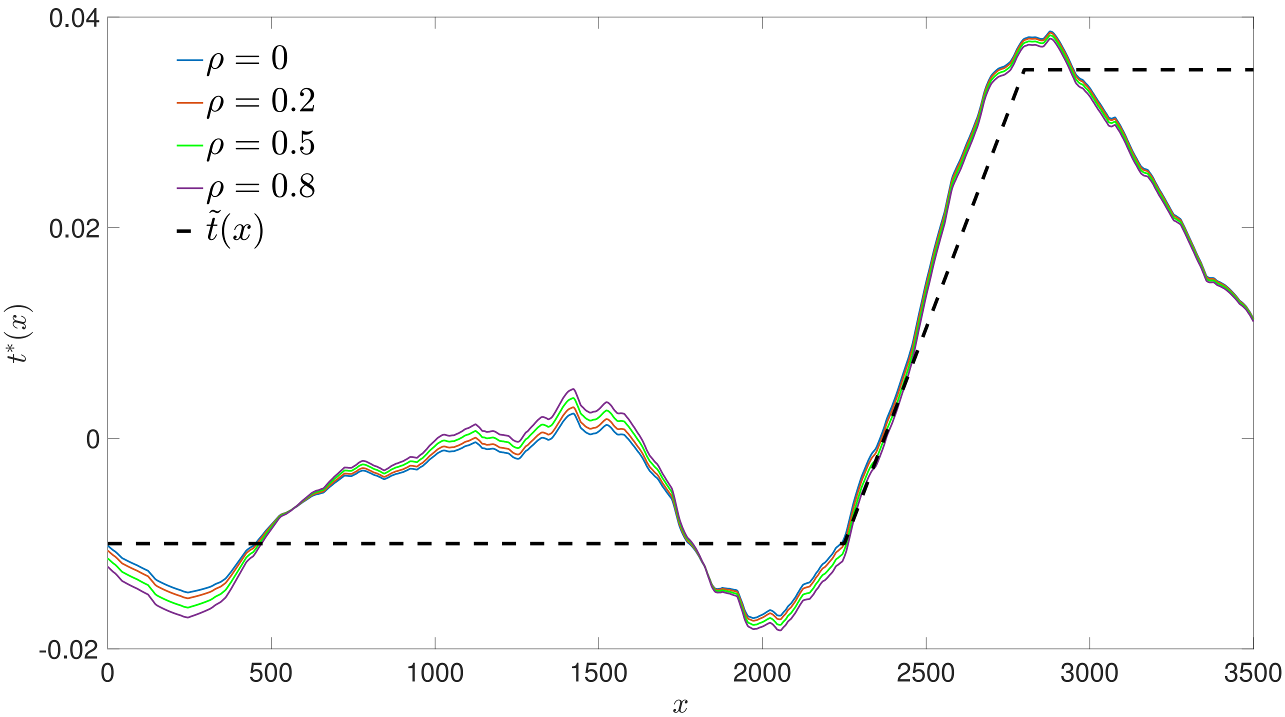

Figure 4: Optimal Perturbation to the status quo contract

regards to extrapolating using local information. For example, she might choose a θ that she expects, based on (7), to lead to a change in effort that is of the same magnitude as $\mathcal{D}a(\widehat{w}, \widehat{t})$. Figure 5 illustrates the new contract, $\widetilde{w}$, as well as $\widehat{w}$ and $\widehat{w} + \widehat{t}$. For this choice of θ, the contract $\widetilde{w}$ is projected to increase the principal's profits by $5-6\%$ depending on the choice of the agent's degree of risk aversion, $\rho \in [0, 1]$. If the principal is willing to extrapolate further and set $\theta = 2000$, then she can (potentially) increase her profit by $9 - 10\%$. We remark that for all values of $\rho \in [0, 1]$, the perturbation $\widetilde{t}$ gives the agent at least as much utility as the status quo contract, $\widehat{w}$. Finally, let us discuss the role of the parameter p. Figure 6 illustrates the optimal perturbation, $\theta_p t^*$, for three different values of p, where, to make them comparable, we set θ_p such that each perturbation is projected to lead to the same change in effort. Observe that if $p = 2$, then the perturbation tailors changes in pay closely to the information, that is, t^* is proportional to $\lambda + \mu \widehat{f}_a / \widehat{f}$. On the other hand, as p becomes larger, the optimal perturbation focuses on the sign of $\lambda + \mu \widehat{f}_a / \widehat{f}$: if it is positive (negative), then it increases (decreases) payments by roughly the same amount. As $p \to \infty$, then t^* equals 1 (resp. -1) if $\lambda + \mu \widehat{f}_a / \widehat{f}$ is positive (resp. negative).

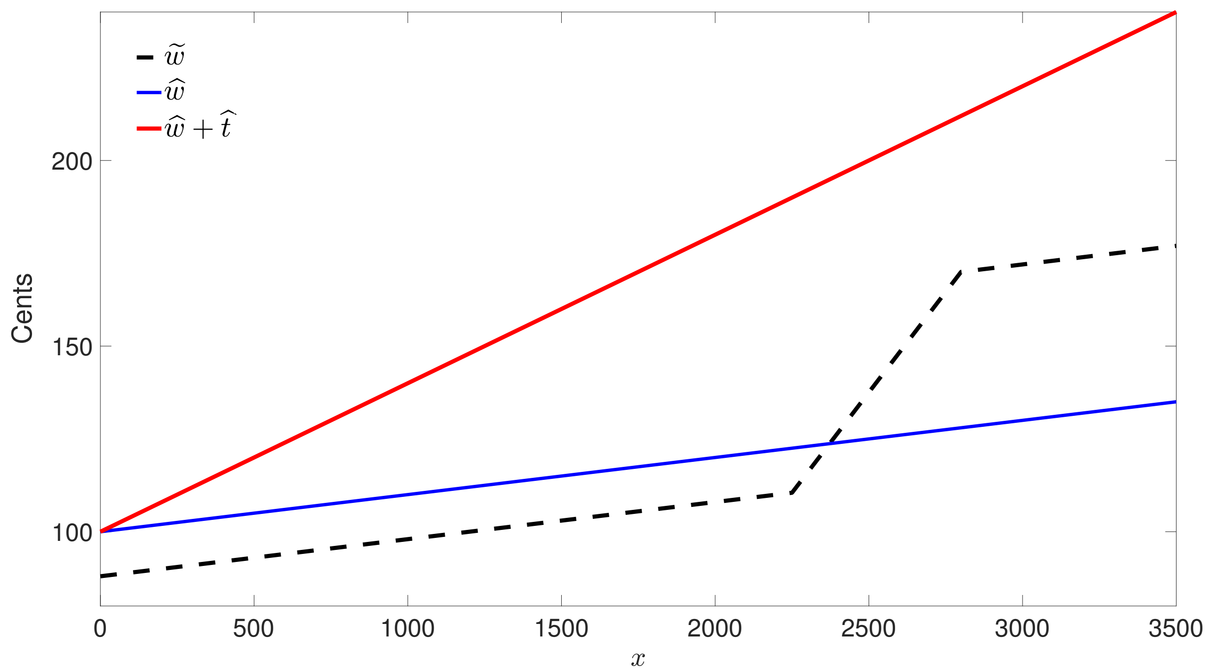

Figure 5: Illustration of the perturbed contract $\widetilde{w}$, the status quo contract $\widehat{w}$, and $\widehat{w}+\widehat{t}$

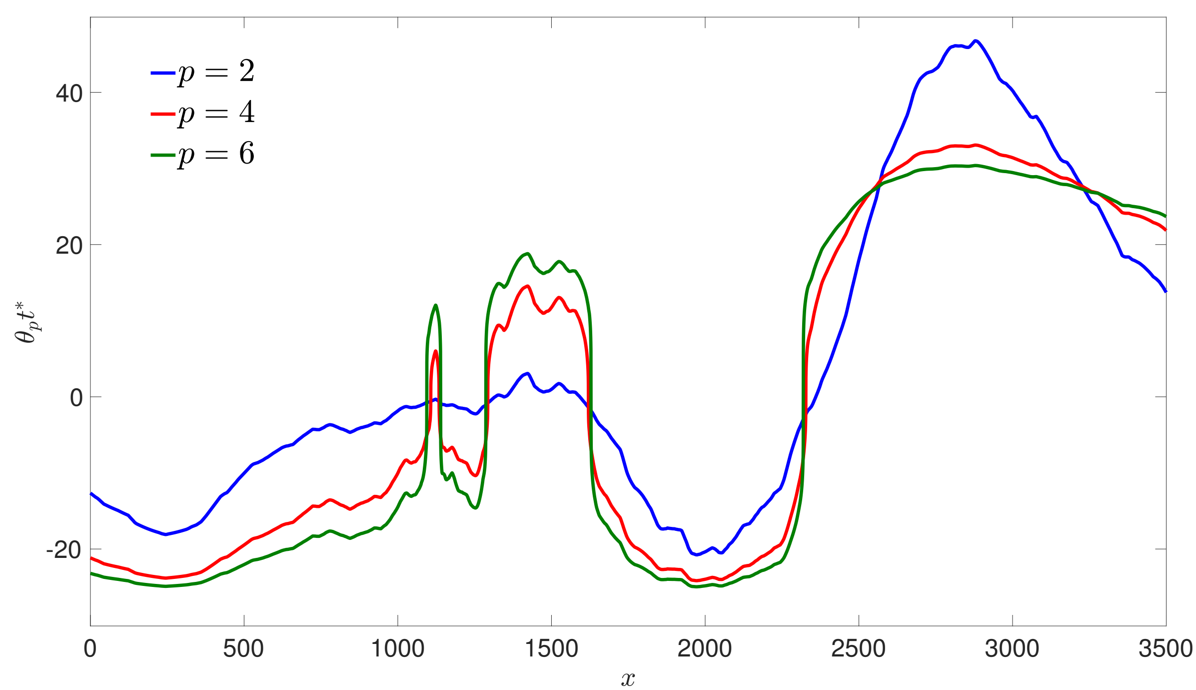

Figure 6: Optimal perturbation, $\theta_p t^*$, for different values of p, where the step size θ_p is chosen such that each perturbation is expected to induce the same change in effort.

7 Discussion

Consider a manager who is interested in optimizing the performance pay plan that she offers to her employee(s). To address this question, we consider a canonical principal-agent framework a-la-Holmström (1979). It is well-known that to characterize the optimal contract, the principal must know all payoff-relevant parameters; specifically, the agent's preferences, his cost function, the distribution of output corresponding to every effort level, and his outside option. In practice, managers rarely have such complete knowledge. As a second best, a manager might look for the *optimal way* to perturb the incentive contract that she currently has in place. To address this question, we assume that the principal has an estimate for the distribution of output corresponding to this *status quo* contract, and, first, we characterize the perturbation that increases her profit at the fastest rate, and second, we examine what information she needs to make that determination.

Our main result shows that if the principal takes a stance on the agent's marginal utility function and she has an estimate for the distribution of output corresponding to some perturbation of the status quo contract, then she can estimate how the agent would respond to *any* other perturbed contract. In turn, this information enables her to determine how to locally improve upon the status quo contract optimally. The same informational requirements hold, and an analogous sufficient statistic result is obtained, when the principal restricts herself to choosing from a lower-dimensional parametric class of contracts; *e.g.*, linear contracts.

References

Baker, G., 2000. The use of performance measures in incentive contracting. *American Economic Review*, 90(2), pp.415-420.

Balbuzanov, I., Gars, J. and Tjernstrom, E., 2017. Media and Motivation: The Effect of Performance Pay on Writers and Content. Working Paper.

Bandiera, O., Barankay, I. and Rasul, I., 2007. Incentives for managers and inequality among workers: Evidence from a firm-level experiment. *Quarterly Journal of Economics*, 122(2), pp.729-773.

Bandiera, O., Barankay, I. and Rasul, I., 2009. Social connections and incentives in the workplace: Evidence from personnel data. *Econometrica*, 77(4), pp.1047-1094.

Barron, D., Georgiadis, G. and Swinkels, J., 2019. Optimal Contracts with a Risk-Taking Agent. Working Paper.

Barseghyan, L., Molinari, F., O'Donoghue, T. and Teitelbaum, J.C., 2018. Estimating risk preferences in the field. *Journal of Economic Literature*, 56(2), pp.501-64.

Bolton, P. and Dewatripont, M., 2005. Contract Theory. MIT Press.

Boyd, S. and Vandenberghe, L., 2004. Convex optimization. Cambridge university press.

Brewer, M., Saez, E. and Shephard, A., 2010. Means-testing and Tax Rates on Earnings. In *Dimensions of Tax Design: The Mirrlees Review.*

Carroll, G., 2015. Robustness and linear contracts. *American Economic Review*, 105(2), pp. 536-563.

Chiappori, P.A. and Salanié, B. 2003. Testing Contract Theory: A Survey of Some Recent Work, in M. Dewatripont, L. Hansen and S. Turnovsky, eds., Advances in Economics and Econometrics, Cambridge University Press, Cambridge.

DellaVigna, S. and Pope, D., 2017. What motivates effort? Evidence and expert forecasts. *Review of Economic Studies*, 85(2), pp.1029-1069.

Diamond, P.A., 1998. Optimal Income Taxation: An Example with a U-shaped Pattern of Optimal Marginal Tax Rates. *American Economic Review*, pp.83-95.

Fehr, E. and Goette, L., 2007. Do workers work more if wages are high? Evidence from a randomized field experiment. *American Economic Review*, 97(1), pp.298-317.

Gayle, G.L. and Miller, R.A., 2009. Has moral hazard become a more important factor in managerial compensation?. *American Economic Review*, 99(5), pp.1740-69.

Garrett, D.F. and Pavan, A., 2015. Dynamic Managerial Compensation: A Variational Approach. *Journal of Economic Theory*, 159, pp.775-818.

Georgiadis, G. and Szentes, B., 2019. Optimal Monitoring Design. Working Paper.

Gibbs, M., 2016. Past, present and future compensation research: Economist perspectives. *Compensation & Benefits Review*, 48(1-2), pp.3-16.

Gibbs, M., Neckermann, S. and Siemroth, C., 2017. A field experiment in motivating employee ideas. *Review of Economics and Statistics*, 99(4), pp.577-590.

Grossman, S. and Hart, O.D., 1983. An Analysis of the Principal-Agent Problem. *Econometrica,* 51(1), pp.7-45.

Guiteras, R.P. and Jack, B.K., 2018. Productivity in piece-rate labor markets: Evidence from rural Malawi. *Journal of Development Economics*, 131, pp.42-61.

Hall, B.J., Lazear, E., and Madigan, C., 2000. Performance Pay at Safelite Auto Glass (A). Harvard Business School Case 800-291.

Hansen, B.E., 2009. Lecture notes on nonparametrics.

Herweg, F., Müller, D. and Weinschenk, P., 2010. Binary payment schemes: Moral hazard and loss aversion. *American Economic Review*, 100(5), pp.2451-77.

Hill, A.E., 2019. The Minimum Wage and Productivity: A Case Study of California Strawberry Pickers. Working Paper.

Holmström, B., 1979. Moral Hazard and Observability. *Bell Journal of Economics*, pp.74-91.

Holmström, B., 2017. Pay for performance and beyond. *American Economic Review*, 107(7), pp.1753-77.

Holmström, B. and Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, pp. 303-328.

Holmström, B. and Milgrom, P., 1991. Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design. *Journal of Law, Economics, & Organization*, 7, pp.24-52.

Hong, F., Hossain, T., List, J.A. and Tanaka, M., 2018. Testing the theory of multitasking: Evidence from a natural field experiment in Chinese factories. *International Economic Review*, 59(2), pp.511-536.

Jensen, M.C., 2001. Corporate Budgeting Is Broken, Let's Fix It. *Harvard Business Review*, 79(10), pp.94-101.

Jewitt, I., 1988. Justifying the First-order Approach to Principal-agent Problems. *Econometrica*, pp. 1177-1190.

Jewitt, I., Kadan, O. and Swinkels, J.M., 2008. Moral Hazard with Bounded Payments. *Journal of Economic Theory*, 143(1), pp.59-82.

Koller, F., 2010. Spark: How Old-Fashioned Values Drive a Twenty-First-Century Corporation: Lessons from Lincoln Electric's Unique Guaranteed Employment Program. PublicAffairs.

Lazear, E.P., 2000. Performance pay and productivity. *American Economic Review*, 90(5), pp.1346-1361.

Mirrlees, J.A., 1976. The Optimal Structure of Incentives and Authority within an Organization. *Bell Journal of Economics*, pp. 105-131.

Oettinger, G.S., 2001. Do piece rates influence effort choices? Evidence from stadium vendors. *Economics Letters*, 73(1), pp.117-123.

Oyer, P. and Schaefer, S., 2011. Personnel Economics: Hiring and Incentives. Volume 4, Part B, Chapter 20 of Handbook of Labor Economics.

Paarsch, H.J. and Shearer, B.S., 1999. The response of worker effort to piece rates: evidence from the British Columbia tree-planting industry. *Journal of Human Resources*, pp.643-667.

Prendergast, C., 2014. The empirical content of pay-for-performance. *Journal of Law, Economics, & Organization*, 31(2), pp.242-261.

Saez, E., 2001. Using Elasticities to derive Optimal Income Tax Rates. *Review of Economic Studies*, 68(1), pp.205-229.

Shearer, B., 2004. Piece rates, fixed wages and incentives: Evidence from a field experiment. *Review of Economic Studies*, 71(2), pp.513-534.

Tirole, J., 1988. The Theory of Industrial Organization. MIT Press.

Wilson, R.B., 1993. Nonlinear pricing. Oxford University Press.

A Robustness

A.1 Other Sources of Incentives

Employees are motivated not only by performance pay, but also by other factors, such as the prospect of a promotion, the threat of firing, prestige, and so on. To capture such indirect incentives, suppose that faced with a contract w, the agent chooses his effort $a(w)$ by solving

$$\int v(w(x)) f_a(x|a(w))dx + I(a(w)) = c'(a(w))\,,$$

for some continuously differentiable function $I(\cdot)$, which captures his marginal benefit from exerting effort due factors other than performance pay. Given the status quo contract $\widehat{w}$ and a perturbation t, the Gateaux derivative of the agent's effort when $\widehat{w}$ is perturbed in the direction of t is

$$\mathcal{D}a(\widehat{w}, t) = \frac{\int t v'(\widehat{w}) \widehat{f}_a dx}{c''(a(\widehat{w})) - \int v(\widehat{w}) f_{aa}(x|a(\widehat{w}))dx - I'(a(\widehat{w}))}\,.$$

Similar to (3), $\mathcal{D}a(\widehat{w}, t)$ is linear in t, and t appears only in the numerator. Therefore, given an estimate of $\mathcal{D}a(\widehat{w}, \widehat{t})$ for some perturbation $\widehat{t}$, the principal can compute the agent's response $\mathcal{D}a(\widehat{w}, t)$ corresponding to any other t using (7). Since $I(\cdot)$ does not enter the principal's profit functions, it is straightforward to verify that the principal's problem is the same as in (P'), and hence Proposition 1 holds verbatim.

Notice that to determine the optimal perturbation, the principal does not need to take a stance on $I(\cdot)$. This is because the agent's marginal benefit of effort is additively separable in direct and indirect incentives, and the function $I(\cdot)$ does not depend directly on the offered contract. Finally, we remark that the assumption that $I(\cdot)$ depends on the agent's effort but not on the realization of x or the agent's utility function is without loss.

A.2 Multiplicatively Separable Utility

In this section, we consider the case in which the agent's preferences are multiplicatively separable in money and effort, that is, we assume that if the agent is paid ω and exerts effort a, then he obtains utility $v(\omega)c(a)$ where v and c satisfy the usual assumptions and $v < 0$. This analysis allows us to capture the case in which the agent's utility exhibits constant relative risk aversion, and hence $v(\omega) = -\exp(-\lambda\omega)$ for some $\lambda > 0$.

Suppose that the first-order approach is valid, so that given contract w, the agent's optimal effort, $a(w)$, satisfies

$$\int v(w(x))\left[c'(a(w))f(x|a(w)) + c(a(w))f_a(x|a(w))\right]dx = 0\,. \tag{25}$$

Using (25), the Gateaux derivative of the agent's effort when the contract $\widehat{w}$ is perturbed in the direction of some t is equal to

$$\mathcal{D}a(\widehat{w},t) = \frac{-\int tv'(\widehat{w})\left[c'(a(\widehat{w}))\widehat{f} + c(a(\widehat{w}))\widehat{f}_a\right]dx}{\int v(\widehat{w})\left[c''(a(\widehat{w}))\widehat{f} + 2c'(a(\widehat{w}))\widehat{f}_a + c(a(\widehat{w}))f_{aa}(x|a(\widehat{w}))\right]dx}\,.$$

Similar to the additively separable case analyzed in Section 3.1, $\mathcal{D}a(\widehat{w},t)$ is linear in t, and t appears only in the numerator. Thus, the agent's response following any perturbation t can be expressed as a function of $\mathcal{D}a(\widehat{w},\widehat{t})$ as follows:

$$\mathcal{D}a(\widehat{w},t) = \frac{\mathcal{D}a(\widehat{w},\widehat{t})}{\int \widehat{t}v'(\widehat{w})\left[\frac{c'(a(\widehat{w}))}{c(a(\widehat{w}))}\widehat{f} + \widehat{f}_a\right]dx}\int tv'(\widehat{w})\left[\frac{c'(a(\widehat{w}))}{c(a(\widehat{w}))}\widehat{f} + \widehat{f}_a\right]dx\,. \tag{26}$$

To evaluate (26), the principal must know $\widehat{f}$ and $\widehat{f}_a$, which she does by assumption, she needs to take a stance on the agent's marginal utility function v', and additionally, she needs to know $c'(a(\widehat{w}))/c(a(\widehat{w}))$. Notice that this ratio can be inferred from (25) if the principal takes a stance on the agent's utility function v (instead of the marginal, v') by computing

$$\frac{c'(a(\widehat{w}))}{c(a(\widehat{w}))} = -\frac{\int v(\widehat{w})\widehat{f}_a dx}{\int v(\widehat{w})\widehat{f}dx}\,.$$

It is straightforward to verify that the expression for $\mathcal{D}\pi(\widehat{w},t)$ and the constraint that the principal restricts attention to perturbations which give the agent at least as much utility as the status quo contract are the same as in the case with additively separable utility, given in (5) and (6), respectively. Therefore, letting

$$\mu := \frac{\left(m - \int \widehat{w}\widehat{f}_a dx\right)\mathcal{D}a(\widehat{w},\widehat{t})}{\int \widehat{t}v'(\widehat{w})\left[\frac{c'(a(\widehat{w}))}{c(a(\widehat{w}))}\widehat{f} + \widehat{f}_a\right]dx}\,,$$

the principal's problem can be expressed as

$$\begin{aligned} \max_{t \text{ u.s.c}} \; & \mu \int t v'(\widehat{w}) \left[\frac{c'(a(\widehat{w}))}{c(a(\widehat{w}))} \widehat{f} + \widehat{f}_a \right] dx - \int t \widehat{f} dx \\ \text{s.t.} \; & \int t v'(\widehat{w}) \widehat{f} dx \geq 0 \\ & \int |t|^p \, dx \leq 1 \end{aligned} \tag{P''}$$

This is a convex optimization program, and it can be solved using the same approach that we used to solve (P') in Section 4.1. We omit the details and conclude with the following remark.

Remark 2. *Assume that if the agent is paid ω and exerts effort a, then he obtains utility $v(\omega)c(a)$ where v and c satisfy the usual assumptions and $v < 0$. If the principal knows v, then she can evaluate whether the status quo contract $\widehat{w}$ is locally optimal, and if not, compute the perturbation which solves (P”).*

B Proofs

Proof of Proposition 1. First, let us characterize the solution to (P'). Recall that the dual program is convex (even if the primal is not convex), because it is the pointwise minimum of affine functions. Therefore, the multipliers λ^* and ν^* obtained in (10) and (11) are necessary and sufficient for an optimum in the dual program.

We will now show that strong duality holds. Towards this goal, let Π^* denote the optimal value of the primal program given in (P'). Weak duality implies that $L(\lambda^*, \nu^*) \geq \Pi^*$. Moreover, it is straightforward to verify that $t(\lambda^*, \nu^*)$ is feasible for (P'), and λ^* and ν^* is strictly positive if and only if the respective (primal) constraint binds. This implies that the objective of (P') evaluated at $t(\lambda^*, \nu^*)$ is equal to $L(\lambda^*, \nu^*)$, and it must be the case that $L(\lambda^*, \nu^*) \leq \Pi^*$. Therefore, we conclude that $L(\lambda^*, \nu^*) = \Pi^*$, which proves that the perturbation $t(\lambda^*, \nu^*)$ is optimal for (P').

To complete the proof, we show that $\widehat{w}$ is locally optimal if and only if (12) is satisfied for all x. Clearly, $\widehat{w}$ is locally optimal if and only if the optimal perturbation $t^* = 0$ for all x, which is true only if for some $\lambda' \geq 0$, we have $\left(\lambda' \widehat{f} + \mu \widehat{f}_a\right) v'(\widehat{w}) = \widehat{f}$ for every x. Suppose that this is the same (for some $\lambda' \geq 0$). Integrating both sides with respect to x and using that $\int \widehat{f}_a dx = 0$ implies that $\lambda' = \int \widehat{f}/v'(\widehat{w}) dx$. It is straightforward to verify that $t \equiv 0$ solves (9) when $\lambda = \lambda'$, and $L(\lambda', \nu) = \nu$ for every ν. Therefore, $\min_{\nu \geq 0} L(\lambda', \nu) = 0$, and weak duality implies that the value of the primal program is bounded by 0 from above. As

$t \equiv 0$ is feasible for the primal, and the objective equals 0 when $t \equiv 0$, it follows that $t \equiv 0$ is indeed the optimal perturbation. $\square$

Proof of Proposition 2. See Proposition 1 in Jewitt, Kadan, and Swinkels (2008). $\square$

Proof of Proposition 3. To begin, fix the dual multipliers associated with the first and second constraint of (23), denoted by λ and ν, respectively. The Lagrangian is

$$L(\lambda, \nu) = \max_{z \in \mathbb{R}^n} \nu + \sum_{i=1}^{n} \left\{ z_i \left[\mu_{\widehat{z}} \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f}_a dx - \frac{\partial}{\partial \alpha_i} \int w_{\widehat{\alpha}} \widehat{f} dx + \lambda \frac{\partial}{\partial \alpha_i} \int v(w_{\widehat{\alpha}}) \widehat{f} dx \right] - \nu z^2 \right\} .$$

It is straightforward to verify that the value of z_i which minimizes the Lagrangian satisfies (24). By substituting this value in the Lagrangian, noting that the dual program is convex, and solving

$$\min_{\lambda \geq 0 \,,\, \nu \geq 0} L(\lambda, \nu) \,, \tag{27}$$

it is straightforward to verify that the optimal multipliers satisfy the expressions given in Proposition 3. Finally, because $z \equiv 0$ is feasible and the primal, (23), is convex, we can use Slater's condition to conclude that strong duality holds, and so the solution (27) is a solution to (23). $\square$