Robust Contracts: A Revealed Preference Approach*

Nemanja Antic and George Georgiadis†

April 2, 2024

Abstract

We study an agency model in which the principal knows the agent-optimal actions in response to K "known" contracts but is unaware of other actions available or their costs, and seeks a contract to maximize worst-case profits. The optimal contract is a mixture of the known contracts and output. Moreover, when $K = 1$, the single known contract maximizes the principal's profit guarantee, whereas with two known contracts, the optimal mixture puts positive weight on one of the known contracts. Our methodology is straightforward to implement, a point that we demonstrate using data from an experimental study of different incentive schemes.

*We are grateful to Alessandro Bonatti, Paul Duetting, Peter Klibanoff, Michael Powell, to participants at various seminars and conferences for helpful comments, and to David Almog Salama Finkelstein for excellent research assistance.

†N. Antic: Northwestern University, nemanja.antic@kellogg.northwestern.edu; G. Georgiadis: Northwestern University, g-georgiadis@kellogg.northwestern.edu

1 Introduction

Firms and organizations throughout the economy use performance pay to motivate their employees. Proper design of incentive schemes however is crucial: when Safelite Autoglass switched from hourly wages to piece rates for their key workers, productivity increased by 44% year-to-year (Lazear, 2000). On the other hand, poorly designed incentives can have dire, sometimes even catastrophic consequences (Jensen, 2002 and Rajan, 2011).

To introduce the main ideas and motivate some of our modeling choices, imagine that you run a car dealership and want to design a new incentive plan for your salespeople. To simplify matters, suppose you have settled on rewarding salespeople according to monthly sales, and all that remains to decide is the pay-for-performance relationship. One approach you could take is to adopt industry best practices (see, for example, Zoltners, Sinha and Lorimer, 2006). You could also take guidance from contract theory: make assumptions about the production environment—the employees' action set, how actions map into outcomes, and their preferences over money and actions, and then exploiting variation in the offered incentives to recover the unknown parameters (Misra and Nair, 2011 and Georgiadis and Powell, 2022). Some managers, however, may be uncomfortable making such arguably strong assumptions, perhaps due to a lack of information about the production environment. In this paper, we characterize optimal incentives given outcome data from a set of incentive schemes, but otherwise minimal assumptions about the production environment.

In our principal-agent model, events unfold as follows: First, the principal offers a contract, which specifies a non-negative payment to the agent as a function of realized output. Then the agent chooses a costly action—a probability distribution over output—to maximize his expected payoff. Finally, output is drawn according to the chosen distribution, and payoffs are realized. The principal has outcome data under K different exogenous contracts which, sidestepping estimation error, enables her to recover the action corresponding to each of these contracts. We assume that the agent best-responds to the offered contract and has quasi-linear preferences over money and actions, but we make no further assumptions about the production environment. The principal does not have prior beliefs about any of the unknown aspects of the environment, and she seeks a contract that maximizes the worst-case profit.[1]

We begin with the benchmark case in which there is a single known contract (i.e., $K = 1$). In this case, we show that the known contract provides the largest possible profit guarantee. To see why, suppose the agent has only two possible actions to choose from: the known action which is known to be "productive" and a completely unproductive one. If the principal offers a contract that pays more in expectation under the productive action, the principal's profit clearly decreases relative to the known contract. But, if the principal offers a contract that pays less, then nature will choose the costs of the productive action to be so high that the agent now prefers the unproductive action, harming the principal again.

We then turn to the case with two known contracts. This case is of particular interest considering that firms are notoriously reluctant to experiment with different incentive schemes, and the majority of studies that exploit variation in incentives feature outcome data from two contracts; see for example Lazear (2018) and the references therein. We show that under certain conditions, a mixture of *one* of the known contracts and the linear one that makes the agent residual claimant maximizes the principal's profit guarantee and is therefore optimal. Furthermore, if both known contracts are linear, then these conditions are never met, in which case the more profitable of the known contracts is optimal.

With more than three known contracts, the optimal contract is a convex combination of the known contracts and the linear one that makes the agent residual claimant. In addition, we propose a two-step procedure to obtain the optimal contract numerically. This procedure first fixes a subset of the Lagrange multipliers and solve a linear program; then it finds the optimal multipliers by solving a non-convex program using simulated annealing.

To demonstrate the applicability of our methodology, we use data from DellaVigna and

[1] This is in the spirit of Carroll (2015) where the principal behaves as if, after committing to a contract, an adversarial third-party, "nature", chooses the unknown aspects of the environment—in our case, the actions available to the agent and their respective costs—to minimize her profit (subject to a set of revealed preference constraints).

Pope's 2018 large-scale experimental study of how different incentive schemes motivate subjects in a real-effort task. For every subset of the seven treatments in which subjects were motivated solely by financial incentives, we take that subset to constitute the set of "known" contracts. We then use the outcome data from each treatment in that subset to compute the empirical distribution function, which corresponds to the agent's optimal action under that treatment, and we compute the optimal contract. In each of the 127 subsets and all values of the marginal value of output that we consider, we find that the most profitable of the known contracts provides the largest profit guarantee; i.e., a mixture contract is never optimal.

Related Literature. Our paper builds on the literature that studies principal-agent problems under moral hazard pioneered by Holmström (1979) and Mirrlees (1976). In particular, we contribute to the strands of this literature that have sought to relax the knowledge assumptions in the canonical model; see Georgiadis (Forthcoming) for an overview. One strand studies models in which the principal is oblivious to one or more parameters and designs a mechanism to elicit this information from the agent (Alon et al., 2023; Chade et al., 2022; Castro-Pires and Moreira, 2021; Gottlieb and Moreira, 2021).

The strand which our model is closest to takes the stance that the principal is ambiguity-averse and pursues "robust" contracts that provide the largest profit guarantee. Carroll (2015) shows that if the principal knows only a subset of the actions available to the agent and their costs, then a linear contract is optimal. Walton and Carroll (2022) provide more general conditions for linear contracts to be optimal. Dütting, Roughgarden and Talgam-Cohen (2019) extend this by showing that linear contracts are max-min optimal in a model where the principal knows, for each action, its expected output but not its distribution. In a setting where the principal also knows a "lower bound" distribution, Antic (2022) shows that optimal contracts are mixtures of debt and equity. See also Carroll (2019) for an overview. Instead, in our model the principal does not know the costs of *any* of the agent's actions but can partially infer them from the agent's revealed preferences. In a concurrent paper, Burkett and Rosenthal (2022) consider, in effect, the same problem. While some of our results are similar (e.g., our Theorem 3 and Corollary 1 parallel their Propositions 3 and 5, respectively), they focus on conditions on the underlying data under which the optimal contract is a mixture of *one* of the known contracts and a linear one. Instead, we fully characterize the optimal contract when $K \in \{1, 2\}$, we propose an optimization algorithm to obtain it numerically when $K \geq 3$, and we demonstrate how our methodology can be applied using data from a real-effort experiment.

2 Model

We consider a contractual relationship between a principal (she) and an agent (he). The principal designs a contract $w : \mathcal{X} \to \mathbb{R}_+$, an upper-semicontinuous mapping from the set of feasible outputs $\mathcal{X} = [0, \overline{x}]$ to non-negative payments to the agent.[2] Then the agent chooses an action F, which is a probability distribution supported on $\mathcal{X}$, by paying a private cost $C(F)$. Output $x \sim F$ is drawn and payoffs are realized.

Let $\mathcal{F} \subseteq \Delta(\mathcal{X})$ denote the agent's action set. The principal does not have full knowledge of this set. However, she knows that it includes a costless action F_0 that generates zero output with certainty, and knows the action that the agent has chosen in response to each of K "known" contracts. We denote these contracts by $w_1, \ldots, w_K$, and the respective actions by $F_1, \ldots, F_K$. Importantly, the principal does not know the cost associated with each of these actions. However, she knows that the agent is rational, and chooses a payoff-maximizing action. An interpretation is that the principal has observational data from having offered each of these contracts enabling her to compute the respective output distribution chosen by the agent.[3]

The agent is risk-neutral, has outside option 0, and is cash-constrained. Therefore,

[2] We restrict attention to deterministic contracts for simplicity and realism.

[3] We assume that these contracts are exogenous and that the agent narrowly best-responds to the offered contract without any strategic considerations.

any feasible contract must specify non-negative payments. The principal has quasilinear preferences and given her knowledge, evaluates each contract according to its worst-case profit. Specifically, this worst-case profit when she offers contract w equals

$$\begin{aligned}
\Pi(w) := \inf_{\mathcal{F},C} & \int [mx - w(x)] dF(x) \\
\text{s.t. } & F \in \arg\max_{\widetilde{F}\in\mathcal{F}} \left\{ \int w(x) d\widetilde{F}(x) - C(\widetilde{F}) \right\} && \text{(IC)} \\
& F_k \in \arg\max_{\widetilde{F}\in\mathcal{F}} \left\{ \int w_k(x) d\widetilde{F}(x) - C(\widetilde{F}) \right\} \text{ for all } k \in \{1, \dots, K\} && \text{(RP)} \\
& \mathcal{F} \supseteq \{F_0, \dots, F_K, F\} \text{ and } C(F) \geq 0 \text{ for all } F \in \mathcal{F} \text{ with } C(F_0) = 0,
\end{aligned}$$

where m is the principal's gross profit per unit of output.[4] This is as-if after the principal offers a contract, an adversarial third-party—*nature*—chooses the agent's action set and the cost of each action to minimize the principal's profit subject to (IC), which specifies the agent's best response to the offered contract, w, and a set of revealed preference constraints given in (RP), which impose that each F_k is a best response to w_k. Naturally the action set $\mathcal{F}$ must include F, as well as the known actions $F_0, \dots, F_K$, and costs must be nonnegative. The principal's objective is to find a contract that maximizes her worst-case profit:

$$\Pi^* = \sup_{w \geq 0} \Pi(w). \tag{P}$$

A contract is *optimal* if it gives profit guarantee Π^* to the principal.

Finally, we impose three assumptions on the K known contract-action pairs:

(**A.1**) Contract w_1 delivers the largest payoff to the principal and $\int [mx - w_1(x)]\, dF_1(x) > 0$.

(**A.2**) Each contract has $w_k(x) \geq 0$ for all $x \in \mathcal{X}$ and $w_k(0) = 0$.

(**A.3**) The agent's best responses can be rationalized; i.e., nature's problem is feasible (see Rochet, 1987).

[4] All integrals are evaluated from $x = 0$ to $x = \overline{x}$. We omit these limits for notational simplicity. When convenient, we will also omit the argument of functions and write, for example, w instead of $w(x)$.

Assumption A.1 ensures that the principal does not prefer to walk away. The first part of A.2 states that the known contracts respect limited liability. The second part ensures that none of the known contracts can be trivially improved by a downward shift until the agent's limited liability constraint binds. Assumption A.3 is necessary for the problem to be feasible; if it fails for some j and k, then no action costs can simultaneously rationalize the agent choosing F_j over F_k when contract w_j is offered, and choosing F_k over F_j when w_k is offered.

2.1 The Principal's Problem Simplified

In this subsection, we show that the principal's problem is equivalent to the following simpler, more tractable formulation (due to nature's problem being a linear program):

$$\sup_{w_{K+1}} \inf_{F_{K+1},\mathbf{c}} \int [mx - w_{K+1}(x)]dF_{K+1}(x) \tag{P$'$}$$

$$\text{s.t.} \int w_k(x)dF_k(x) - c_k \geq \int w_k(x)dF_j(x) - c_j \text{ for all } k \text{ and } j \neq k \tag{IC-RP}$$

$$w_{K+1}(\cdot) \geq 0,\ F_{K+1} \in \Delta(\mathcal{X}),\ \text{and } \mathbf{c} \in \mathbb{R}_+^{K+1},$$

where $k \in \{1, \ldots, K+1\}$ and $j \in \{0, \ldots, K+1\}$. In this formulation, for each contract w_{K+1}, instead of choosing the agent's action set and the cost of each action, nature chooses *one* action, F_{K+1} and the vector $\mathbf{c} = \{c_1, \ldots, c_{K+1}\}$, where c_k is the cost of action F_k, to minimize the principal's profit subject to a set of incentive compatibility and revealed preference constraints, which stipulate that F_k is a best response to w_k for each k. Then the principal chooses w_{K+1} to maximize this worst-case profit.

Lemma 1. *A contract w_{K+1} solves (P) if and only if it solves (P$'$).*

Towards a contradiction, suppose that the action set contains at least two actions beyond the known ones. Since the agent can choose at most one of them in response to the offered contract, nature is no worse off by excluding the additional actions from $\mathcal{F}$. Adding extra actions on the other hand, increases the number of revealed preference constraints, which can only benefit the principal. □

3 Results

In this section we establish our main results. We start off with the case in which there is *one* known contract (i.e., $K = 1$), and show that continuing to offer the same contract maximizes the principal's worst-case profit. Next we characterize the optimal contract for the case with two known contracts. Finally, we consider the case with an arbitrary number of known contracts.

3.1 A Benchmark: One known contract ($K = 1$)

If the principal knows only the agent's best response to a single contract, then she can do no better than continue to offer that same contract.

Theorem 1. *With one known contract, w_1, the principal's worst-case profit is maximized when she offers w_1. In particular, any contract which solves (P') is F_1-a.e. equivalent to w_1.*

For a sketch of the argument, fix an arbitrary contract $w_2 \neq w_1$. If $\int w_1(x)dF_1(x) > \int w_2(x)dF_1(x)$, then nature can induce the agent to choose the null action F_0 by endowing him with no additional actions (e.g., by setting $F_2 \equiv F_0$) and making action F_1 sufficiently costly. Since F_0 results in non-positive profit, the principal prefers to offer w_1. If the inequality is reversed, then nature can induce the agent to choose F_1 in response to w_2 (by endowing him with no additional actions), in which case the principal is again better off offering w_1. Finally, if the inequality binds, then nature can ensure that the principal earns a vanishingly small worst-case profit, which is strictly smaller than the strictly positive one provided by w_1.[5]

[5]While we implicitly assume the agent breaks ties in a way that hurts the principal, in the proof, the agent is never indifferent between actions in his choice set.

This result contrasts with much of the robust contracting literature, which shows that linear contracts are optimal (Carroll, 2015; Dai and Toikka, 2022; Dütting, Roughgarden and Talgam-Cohen, 2019; Walton and Carroll, 2022). The key difference is that this literature assumes that a subset of the agent's actions and their costs are known, and so the first part of the above sketch breaks down: if $\int(w_1(x) - w_2(x))dF_1(x)$ is sufficiently small, then nature may not be able to induce the agent to choose F_0.

Revealed preference constraints bound the costs of known actions. When the principal knows the agent's optimal action in response to only one contract, this bound is too weak for her to improve her payoff. Tighter bounds may arise if, for example, it is common knowledge that the agent earns sufficiently large rents from the observed contract. In this instance, a linear contract is optimal—in line with most of the robust contracting literature (e.g., Carroll, 2015); see Appendix B.1 for details. Tighter bounds on costs also arise when the principal observes the agent's optimal action under multiple contracts. Revealed preference constraints then put stronger restrictions on what costs can be assigned to each action. We examine this case next.

3.2 Two known contracts ($K = 2$)

In this section we suppose that the principal knows the agent-optimal action in response to each of *two* contracts. This case is empirically relevant because most studies that examine the effects of incentives exploit variation from exactly two incentive schemes; see for example Lazear (2018) and the references therein.

To simplify the exposition we introduce some notation. For each i and j, define

$$v_{ij} := \int w_i(x)dF_j(x) \quad \text{and} \quad \mu_j := \int x dF_j(x)$$

to denote the expected payment under w_i if the agent chooses action F_j, and the expected output under this action, respectively. Next, we define

$$\phi := v_{11} + v_{22} - v_{12} - v_{21}.$$

This quantity is non-negative by Assumption A.3, and it relates to the "wiggle room" nature has to hurt the principal by varying c_1 and c_2 while respecting the agent's revealed preference constraints.[6] Finally, for each j and $i \neq j$, and conditional on $m\mu_i - v_{ji} \geq \phi$, define the contract

$$w_j^*(x) := \rho_j w_j(x) + (1 - \rho_j)mx, \text{ where } \rho_j := 1 - \sqrt{\phi/(m\mu_i - v_{ji})},$$

which is a mixture of w_j and the linear contract that makes the agent residual claimant, mx. The following theorem shows that under certain conditions, one of these mixture contracts is optimal; otherwise w_1, the more profitable of the known contracts, is optimal.

Theorem 2. *Suppose the principal knows the contract-action pairs* (w_1, F_1) *and* (w_2, F_2).

(i). If $\sqrt{m\mu_2 - v_{12}} - \sqrt{\phi} > \sqrt{m\mu_1 - v_{11}}$, *then* w_1^* *is optimal;*

(ii). If $\sqrt{m\mu_1 - v_{21}} - \sqrt{\phi} > \sqrt{m\mu_1 - v_{11}}$, *then* w_2^* *is optimal;*

(iii). Otherwise, the more profitable of the known contracts, w_1, *is optimal.*

These conditions are mutually exclusive. The left-hand side of the first and the second condition is the square root of the principal's profit when she offers w_1^* and w_2^*, respectively, while the right-hand sides are the square root of her profit when she offers w_1.

To interpret condition (i), suppose that the principal could offer w_1 and get the agent to choose F_2 instead of F_1; she would benefit if $m\mu_2 - v_{12} > m\mu_1 - v_{11}$. Of course, the principal cannot achieve this aim simply by offering w_1, because it violates one of the revealed preference constraints. Instead, she must appropriately modify incentives, and ϕ relates to the profit she must give up to do so. The interpretation of condition (ii) is analogous.

[6] The revealed preference constraints (IC-RP) stipulate $v_{11} - c_1 \geq v_{12} - c_2$ and $v_{22} - c_2 \geq v_{21} - c_1$, which can be rewritten as $v_{21} - v_{22} \leq c_1 - c_2 \leq v_{11} - v_{12}$. If $\phi = 0$, then $c_1 - c_2$ is pinned down by these constraints, and the larger ϕ is, the more flexibly nature can choose the costs of the known actions.

We make three remarks. First, the mixture contract described above can be implemented by adding an equity award while scaling down the original incentive plan proportionally. Second, conditions (i) and (ii) are easier to satisfy when ϕ is small; in this case however, the optimal contract assigns little weight on mx, so it similar to w_1 or w_2, respectively. And finally, notice that if the known contracts $w_1(x), w_2(x) \leq mx$ for all x, that is, we have limited liability for the principal, w_i^* is Pareto-improving relative to w_i.

Which action does the agent choose in response to the optimal contract? If w_1 is optimal, then of course, the agent chooses F_1. If the principal optimally offers w_j^*, then it can be shown that nature best-responds by endowing the agent with the action

$$F_j^*(x) = \rho_j F_i(x) + (1 - \rho_j) F_0(x),$$

where $i \neq j$, and moreover, the cost of this action is (weakly) smaller than $\rho_j c_i$; see Proposition 1 in Appendix A.4 for details.[7] That is, in response to w_1^*, nature endows the agent with an action that is a mixture of F_2 and F_0 with respective weights ρ_1 and $1 - \rho_1$ (and analogously for w_2^*).

Linear incentive schemes are common in practice. However, if both known contracts are linear, then it is straightforward to verify that conditions (i) and (ii) of Theorem 2 can never satisfied, and so w_1 is optimal.

Corollary 1. *Suppose that both known contracts, w_1 and w_2, are linear. Then w_1 is optimal.*

If w_1, w_2 are linear, by Theorem 2, any third contract is also linear. To see that a new contract cannot improve the principal's profit guarantee, assume, for example, that w_1 offers higher equity to the agent (besides a higher profit for the principal). The third contract always offers more equity than w_2. If it offers more equity than w_1, then nature can endow the agent only with the observed actions (F_0, F_1 and F_2). Any costs consistent with revealed

[7] Recall that nature can be interpreted as an adversarial third-party who chooses the actions available to the agent and their respective costs to minimize the principal's payoff.

preference result in the agent choosing F_1, which is strictly worse for the principal than offering w_1. If the new contract offers equity in-between w_1 and w_2, then nature can choose the costs so that the agent chooses F_2 under the new contract. The new contract is then less attractive than w_2 and hence w_1.[8]

This result suggests that for an ambiguity-averse principal, experimenting only with linear contracts may be counterproductive; instead, it is valuable to also have outcome data under nonlinear contracts.

3.3 K known contracts

In this section we extend our analysis to an arbitrary number of known contracts. Characterizing the optimal contract in this case is challenging because it involves solving a non-convex optimization program. Nevertheless, we can show that it is a convex combination of the known contracts and mx. Moreover, we propose an optimization procedure to solve for the optimal contract, which we use in our empirical exercise in Section 4.

Consider the following maximization program, which is the dual of (P′):

$$\begin{aligned}
\sup_{\boldsymbol{\lambda}\in\mathbb{R}_+^{(K+1)\times(K+2)}} \quad & \frac{\sum_{j=1}^{K}\lambda_{K+1,j}\left(m\mu_j+\sum_{k=1}^{K}\lambda_{k,K+1}v_{kj}\right)}{1+\sum_{j=0}^{K}\lambda_{K+1,j}} \\
& \qquad -\sum_{k=1}^{K}\left(\lambda_{k,K+1}+\lambda_{k0}\right)v_{kk}+\sum_{k=1}^{K}\sum_{j=1}^{K}\lambda_{kj}(v_{kj}-v_{kk}) \\
\text{s.t.} \quad & \lambda_{k,K+1}+\lambda_{k0}+\sum_{j=1}^{K}(\lambda_{kj}-\lambda_{jk})\geq\lambda_{K+1,k} \quad \text{for all } k\in\{1,\ldots,K\} \\
& \sum_{k=1}^{K}\lambda_{k,K+1}\leq\sum_{j=0}^{K}\lambda_{K+1,j}
\end{aligned} \tag{D′}$$

Each λ_{kj} represents the Lagrange multiplier associated with (IC-RP$_{kj}$), which stipulates that

[8] This is reminiscent of the observation by Dütting, Feldman and Peretz (2023) that linear contracts are non-manipulable, in the sense that a principal cannot gain from employing an ambiguous contract when restricted to only using linear contracts.

when offered contract w_k, the agent prefers action F_k to F_j. The following theorem shows that, first, every optimal contract is a convex combination of the K known contracts and the (linear) one that makes the agent residual claimant, and second, the principal's problem is equivalent to (D′).

Theorem 3. *Given K known contracts, a contract w_{K+1} is optimal if and only if it solves (D′). Moreover, every optimal contract takes the form*

$$w_{K+1}(x) := \sum_{k=1}^{K} \rho_k w_k(x) + \left(1 - \sum_{k=1}^{K} \rho_k\right) mx,$$

where $\rho_k = \lambda_{k,K+1}/(1 + \sum_{j=0}^{K} \lambda_{K+1,j}) \geq 0$ for each k.

Observe that (D′) is non-convex owing to the first term in the objective. As a result, standard optimization methods are generally not guaranteed to yield a global maximum. Towards a practical procedure to solve this program, notice that if we fix the multipliers $\{\lambda_{K+1,j}\}_{j=0}^{K} =: \boldsymbol{\lambda}^{K+1}$, then (D′) reduces to a linear program, which can be solved exactly using standard solvers. Denote the objective evaluated at the optimum of this linear program by $\widetilde{\Pi}(\boldsymbol{\lambda}^{K+1})$. Then it remains to solve

$$\sup\ \widetilde{\Pi}(\boldsymbol{\lambda}^{K+1}) \text{ subject to } \boldsymbol{\lambda}^{K+1} \in \mathbb{R}_{+}^{K+1}.$$

While this program is also not convex, its dimension is $K + 1$, whereas (D′) has dimension $(K + 1)^2$. Practically, it can be solved relatively swiftly, for example, using a simulated annealing algorithm.[9]

[9] In light of Theorem 2 one might ask whether the optimal contract always puts positive weight on *one* of the known contracts. It turns out that there do exist instances where the optimal contract puts weight on multiple known contracts as Example 2 of Burkett and Rosenthal (2022) demonstrates.

4 Application

In this section we demonstrate the applicability of our methodology using data from DellaVigna and Pope's 2018 real-effort experiment conducted on Amazon's Mechanical Turk. In the experiment, subjects were tasked with repeatedly pressing the 'a' and 'b' keys in alternating order, and received one *point* for every a/b keystroke pair they managed to complete in a ten-minute period.

We focus on the 7 treatments summarized in Table 1, which differ in the monetary incentives offered. Each subject was randomly assigned to a single treatment, received a \$1 participation fee, and performed the task once. In the first treatment, no incentive pay was offered. In treatments 2 to 5, subjects were paid a constant amount per point, whereas in treatments 6 and 7 they received a lump-sum payment (40 and 80 cents, respectively) if they achieved at least 2,000 points.[10] During the course of the treatment, subjects could see the incentive contract they were on, a countdown clock, as well as a running tally of the points accumulated. The dataset includes the number of points achieved by every subject.

Incentive Contract	Avg. #points	Std. Dev.	#Subjects
$\pi^1(x) = 0$	1521	726	540
$\pi^2(x) = 0.001x$	1883	664	538
$\pi^3(x) = 0.01x$	2029	649	558
$\pi^4(x) = 0.04x$	2132	626	562
$\pi^5(x) = 0.10x$	2175	578	566
$\pi^6(x) = 40\mathbb{I}_{\{x\geq 2000\}}$	2136	576	545
$\pi^7(x) = 80\mathbb{I}_{\{x\geq 2000\}}$	2188	530	532

Table 1: This table describes seven of the experimental treatments in DellaVigna and Pope (2018) that differed in the monetary incentives offered to the subjects.

We now describe the exercise that we perform. First, we make an assumption about the principal's gross profit margin m. Then, for each subset of treatments $\mathcal{W} \subseteq \{\pi^1, \ldots, \pi^7\}$, we take it to constitute the set of "known" contracts, and letting K denote the cardinality of

[10]To be precise, in treatment 2, they were paid 1 cent per thousand points, and in treatments 3, 4 and 5, they were paid 1, 4, and 10 cents, respectively, for every hundred points. For simplicity, we assume $x \in \mathbb{N}$.

$\mathcal{W}$, we define the K known contracts $w_1, \ldots, w_K$. Next, we use the outcome data from each treatment in this set to compute the corresponding empirical CDF F_k, which we take to be the agents' best response to w_k.[11] Finally, we compute the optimal contract.

First, we consider all *pairs* of treatments, that is, all sets $\mathcal{W}$ with cardinality 2 (of which there are 21).[12] For each pair of treatments and every m between 0.05 and 1 with a grid size of 0.001, we check which of the conditions in Theorem 2 are satisfied to identify the optimal contract. In each of these combinations, the more profitable of the known contracts delivers the largest profit guarantee.[13]

Next, we consider all sets of treatments with cardinality greater than two. To find the optimal contract, we solve (D′) using a simulated annealing algorithm.[14] We repeat this procedure for each set of three or more treatments (of which there are 99) and $m = \{0.05, 0.1, \ldots, 1\}$. Again, in every combination, the more profitable of the known contracts is optimal.

In applying our model to the DellaVigna and Pope (2018) data set, we find that in every instance, an ambiguity-averse principal finds it optimal to offer the most profitable of the known contracts instead of experimenting with a yet-unseen contract.

5 Discussion

We study an agency model under moral hazard in which the principal faces ambiguity about the actions available to the agent and their costs. The principal has outcome data under K "known" incentive schemes and seeks a contract with the largest profit guarantee. We show that if $K = 1$, then the single known contract is optimal. With two known contracts, a

[11] In doing so, we abstract away from statistical error and we ignore unobserved heterogeneity. We discuss these issues in Section 5. We also define the K contracts in such an order that w_1 generates the largest profit in line with Assumption A.1.

[12] If $K = 1$, then by Theorem 1 the single known contract provides the largest profit guarantee.

[13] That this is true when both known contracts are linear follows from Corollary 1. Therefore, it suffices to check only the pairs in which one (or both) of the known contracts is π^6 or π^7. Note that our focusing on $m \geq 0.05$ is to ensure that at least one of the contracts in $\mathcal{W}$ is profitable per Assumption A.1.

[14] Simulated annealing is a stochastic optimization method for approximating the optimal solution of non-convex optimization programs, where gradient descent algorithms may get "trapped" at a local maximum.

mixture of *one* of the known contracts and the linear contract that makes the agent residual claimant is optimal. If $K \geq 3$, then the optimal contract is a convex combination of the known ones and the aforementioned linear contract, and propose an algorithm to obtain it numerically. Finally, we demonstrate the applicability of our approach using data from DellaVigna and Pope's (2018) experimental study of a variety of incentive schemes.

Our results are consistent with the fact that most firms are reluctant to change their employees' incentive schemes: Without additional assumptions about the production environment, which managers may be hesitant to make, it is often impossible to find a new contract with a bigger profit guarantee. This is always the case when there is one "known" incentive scheme, and it is often the case when there are two. Moreover, our simulation exercise suggests that this is true even with more "known" incentive schemes. Our results also imply that one would expect to see path-dependence in the contract design process, where otherwise similar firms (or divisions within the same firm) may settle on substantially different incentive plans.

Unobserved heterogeneity. In practice, one may aggregate outcome data from many agents who are offered the same contract. In that case, faced with the same contract, different agents may choose different actions, so the empirical distribution function computed using aggregate outcome data is a composition of each agent's (unobserved) action. Such unobserved heterogeneity increases nature's *leverage*, making it only more likely that one of the known contracts is optimal.

Risk-aversion. Our model can be readily extended to the case in which the agent's payoff, for any given wage scheme and action, is $\int u(w(x))dF(x) - C(F)$, where u is a known, strictly increasing, concave function. In this case, the optimal contract is a now nonlinear function of a set of dual multipliers and the known contracts. If instead the principal is oblivious to the agent's utility function, then requiring that the optimal contract be robust to this type of ambiguity makes it more likely that one of the known contracts is optimal.

Cost restrictions. We place no restrictions on the cost of each action other than those implied by revealed preference. It may be interesting to incorporate restrictions on costs while stopping short of assuming that the cost function is known or can be estimated using outcome data.[15]

Estimation error. We have assumed that for each known contract, the principal can identify the agent's action, that is, the distribution function over output. In practice of course, outcome data is finite, which gives rise to estimation error. As an example, suppose instead that the principal only knows that the distribution corresponding to each of the known contracts lies in some ε-ball around an estimated distribution. If the principal were to offer, say, w_i, then nature would endow the agent with the profit-minimizing distribution inside that ε-ball (subject to meeting the agent's revealed preference constraints, of course). Because nature does not have this added flexibility when the principal offers a new contract, we conjecture that estimation error makes unseen contracts comparatively more attractive.

[15]For example, one could posit that the cost function $C(F)$ is monotone in first-order stochastic dominance (e.g., Georgiadis, Ravid and Szentes, 2024), or that it comes from parametric family (e.g., f-divergence as in Hébert, 2018) and the principal faces ambiguity over a set of parameters.

A Appendix

A.1 Proof of Theorem 1

Fix an arbitrary contract $w_2 \geq 0$ that satisfies (IC-RP).

Case 1: Suppose $\int w_1(x)dF_1(x) > \int w_2(x)dF_1(x)$. Let $c_1 \in \left(\int w_2(x)dF_1(x), \int w_1(x)dF_1(x)\right)$, $F_2 = F_0$ and $c_2 = 0$, so that

$$\int w_2(x)dF_2(x) - c_2 = 0 > \int w_2(x)dF_1(x) - c_1,$$

where the equality follows by Assumption A.2 and the inequality by the definition of c_1. The principal's payoff is $\int[mx - w_2(x)]dF_2(x) = 0 < \int[mx - w_1(x)]dF_1(x)$ by Assumption A.1.

Case 2: Suppose $\int w_1(x)dF_1(x) \leq \int w_2(x)dF_1(x)$ and that $w_1 \neq w_2$ on some F_1- positive measure set. By monotonicity of the integral, there must be some F_1- positive measure subset $A \subseteq \mathcal{X}$ on which $w_1(x) < w_2(x)$. Take $\hat{x} \in A$ and $\epsilon > 0$ so that $\epsilon m\hat{x} - \epsilon w_2(\hat{x}) < \int[mx - w_1(x)]dF_1(x)$. Such an ϵ exists by Assumption A.1. Let $F_2(x) = 1 - \epsilon + \epsilon\mathbb{I}_{\{x \geq \hat{x}\}}$, i.e., F_2 puts $1 - \epsilon$ mass on $x = 0$ and ϵ mass on $\hat{x}$. Let $c_1 = \int w_1(x)dF_1(x) - \eta$ for some $\eta > 0$ and c_2 satisfy $(1 - \epsilon)w_1(0) + \epsilon w_1(\hat{x}) - \eta < c_2 < (1 - \epsilon)w_2(0) + \epsilon w_2(\hat{x}) - \eta$. Note that $c_2 \geq 0$ for η sufficiently small. Then

$$\int w_1(x)dF_1(x) - c_1 = \eta > (1 - \epsilon)w_1(0) + \epsilon w_1(\hat{x}) - c_2 = \int w_1(x)dF_2(x) - c_2, \text{ and}$$
$$\int w_2(x)dF_2(x) - c_2 = (1 - \epsilon)w_2(0) + \epsilon w_2(\hat{x}) - c_2 > \eta = \int w_2(x)dF_1(x) - c_1;$$

i.e., the agent's incentive constraints hold. The principal's payoff is $\epsilon[m\hat{x} - w_2(\hat{x})]$, which is less than her payoff under w_1 by choice of ϵ. □

A.2 Proof of Theorem 2

We begin by establishing the following lemma.

Lemma 2. *Suppose the principal knows the pairs* (w_1, F_1) *and* (w_2, F_2)*. If a contract other than* w_1 *is optimal, then it is either* w_1^* *or* w_2^**.*

Proof of Lemma 2. Suppose a contract $w_3 \geq 0$ is optimal for the principal.

Adding and subtracting $(v_{31} - v_{32})$ to ϕ, implies

$$\phi = [v_{31} + v_{22} - v_{32} - v_{21}] + [v_{32} + v_{11} - v_{31} - v_{12}] \geq 0. \tag{1}$$

Because this sum is nonnegative (by Assumption A.3), at least one of the terms in the square brackets is nonnegative. Without loss label $i, j \in \{1, 2\}$ with $i \neq j$ such that

$$v_{3i} + v_{jj} - v_{3j} - v_{ji} \geq 0. \tag{2}$$

Claim 1. *If* w_3 *is optimal, then* $m\mu_i - v_{3i} > m\mu_1 - v_{11}$ *and hence* $v_{3i} < v_{ii}$*.*

Proof of Claim 1. Nature can let $F_3 \equiv F_i$, $c_3 = c_i = v_{ji}$, and $c_j = v_{jj}$, so that (IC-RP) holds, i.e., $v_{ii} - c_i \geq v_{ij} - c_j$, $v_{jj} - c_j \geq v_{ji} - c_i$, and $v_{3i} - c_i \geq v_{3j} - c_j$ by (2) and $\phi \geq 0$. If the principal's payoff under w_3 is larger than under w_1, it must be so for the above distributions and costs, so $m\mu_i - v_{3i} > m\mu_1 - v_{11} \geq m\mu_i - v_{ii}$. This implies $v_{3i} < v_{ii}$. □

Claim 2. *If* w_3 *is optimal, then* $v_{3j} + v_{ii} - v_{3i} - v_{ij} < 0$*.*

Proof of Claim 2. Assume by way of contradiction that $v_{3j} + v_{ii} - v_{3i} - v_{ij} \geq 0$.

Case 1: $v_{3j} \geq v_{jj}$.

Nature can set $F_3 \equiv F_j$, $c_3 = c_j = v_{ij}$, and $c_i = v_{ii}$. The agent prefers F_3 over F_i when offered contract w_3, since $v_{3j} - v_{ij} \geq v_{3i} - v_{ii}$. All other (IC-RP) also hold, since $v_{ii} - v_{ii} \geq v_{ij} - v_{ij}$ and $v_{jj} - v_{ij} \geq v_{ji} - v_{ii}$ by $\phi \geq 0$. But $v_{3j} \geq v_{jj}$ implies $m\mu_j - v_{3j} \leq m\mu_j - v_{jj}$, contradicting that w_3 is optimal.

Case 2: $v_{3j} < v_{jj}$.

Nature sets $F_3 \equiv F_0$, $c_3 = 0$, $c_i = v_{3i}$ and $c_j = v_{3j}$. (IC-RP) are satisfied, so when contract w_3 is offered the agent chooses $F_3 = F_0$. The principal's payoff is $-w_3(0) \leq 0 < m\mu_1 - v_{11}$ and so w_3 cannot be optimal. This proves claim Claim 2. □

By definition $v_{i0} = 0$ and let $v_{0i} = \int [0] dF_i(x) = 0$. From Theorem 1 of Rochet (1987), costs that rationalize the revealed preferences exist as long as for every finite cycle $k(0), k(1), ..., k(m+1) = k(0)$ in $0, 1, 2, 3$, we have $\sum_{l=1}^{m} v_{k(l)k(l)} - v_{k(l)k(l+1)} \geq 0$. Because of the zero terms, these constraints reduce to

$$
\begin{aligned}
v_{ii} - v_{ij} + v_{jj} - v_{ji} &\geq 0 \\
v_{33} - v_{3i} + v_{ii} - v_{i3} &\geq 0 \qquad (3) \\
v_{33} - v_{3j} + v_{jj} - v_{j3} &\geq 0 \\
v_{ii} - v_{i3} + v_{33} - v_{3j} + v_{jj} - v_{ji} &\geq 0 \\
v_{ii} - v_{ij} + v_{jj} - v_{j3} + v_{33} - v_{3i} &\geq 0 \qquad (4)
\end{aligned}
$$

Whenever w_3 is optimal, inequalities (3) and (4) imply the others. The first inequality holds since $\phi \geq 0$. By claim 2, $-v_{3j} > v_{ii} - v_{3i} - v_{ij}$, and hence (4) implies the third inequality. Finally, equation (2) implies $-v_{3i} < -v_{3j} + v_{ii} + v_{ij}$ and so (3) implies the fourth inequality.

We have shown that if w_3 increases the principal's payoff (relative to w_1), then (3) and (4) are the relevant constraints to check. Next, we will argue that if there exists a contract that dominates w_1, then w_1^* or w_2^* maximizes the principal's payoff. To see this, consider

$$
\sup_{w \geq 0} \inf_{F \in \Delta(\mathcal{X})} \int [mx - w(x)] dF(x) \qquad \text{(P")}
$$

$$
\text{s.t.} \int [w(x) - w_j(x)] dF(x) \geq \int [w(x) - w_i(x)] dF_i(x) + \int [w_i(x) - w_j(x)] dF_j(x) \qquad (5)
$$

$$
\int [w(x) - w_i(x)] dF(x) \geq \int [w(x) - w_i(x)] dF_i(x) \qquad (6)
$$

That is, we consider the principal's max-min problem subject to the constraints (3) and (4), where we have replaced w_3 with the choice variable w. We will check ex-post that the optimal w satisfies Claim 1 and Claim 2, which are necessary for this contract to increase the principal's payoff vis-a-vis w_1. We show that whenever the solution to (P") satisfies these conditions, then it coincides with either w_1^* or w_2^* almost everywhere.

Claim 3. *If w satisfies Claim 1, then the right-hand side of (5) is strictly positive and the right-hand side of (6) is strictly negative.*

Proof of Claim 3. The second part of this claim is immediate from the Claim 1, since it concludes that $v_{3i} < v_{ii}$. For the first part of the claim, notice that if the right-hand side of (5) is negative, nature can choose $F = F_0$, in which case the principal's payoff will be no greater than zero. But then the principal would be better off offering contract w_1, which provides her with a strictly positive payoff by assumption. Since w will be relevant only if it increases the principal's payoff vis-a-vis w_1, we can henceforth assume that the right-hand side of (5) is strictly positive and the right-hand side of (6) is strictly negative. □

Let us fix an arbitrary $w \geq 0$ and nonnegative dual multipliers λ and ν. We have the Lagrangian

$$\begin{aligned}
\mathcal{L}(\lambda, \nu, w) = \inf_{F \in \Delta(\mathcal{X})} & \int [mx - (1 + \lambda + \nu)w(x) + \lambda w_j(x) + \nu w_i(x)] dF(x) \\
& + \lambda \int [w(x) - w_i(x)] dF_i(x) + \lambda \int [w_i(x) - w_j(x)] dF_j(x) \\
& + \nu \int [w(x) - w_i(x)] dF_i(x) \\
= & \min_x \{mx - (1 + \lambda + \nu)w(x) + \lambda w_j(x) + \nu w_i(x)\} + (\lambda + \nu) \int w(x) dF_i(x) \\
& - \lambda \left[\int w_i(x) dF_i(x) - w_i(x) dF_j(x) + w_j(x) dF_j(x) \right] - \nu \int w_i(x) dF_i(x).
\end{aligned}$$

The first integral is minimized by a degenerate distribution F. By the Lagrange Duality Theorem (Luenberger, 1997, Theorem 1, p. 224) strong duality holds, and therefore, the solution to (P”) equals

$$\sup_{w\geq 0}\sup_{\lambda,\nu\geq 0}\mathcal{L}(\lambda,\nu,w). \tag{7}$$

Changing the order of maximization, we fix arbitrary multipliers $\lambda,\nu\geq 0$ and consider $\sup_{w\geq 0}\mathcal{L}(\lambda,\nu,w)$. For each x, a marginal increase in $w(x)$ increases the objective at rate $-(1+\lambda+\nu)+(\lambda+\nu)dF_i(x)<0$ if the expression inside the curly brackets is minimized at that particular x, and at rate $(\lambda+\nu)dF_i(x)\geq 0$ otherwise. Therefore, it is without loss to raise $w(x)$ until it (just) minimizes the expression in the curly brackets, and so that expression must be constant in x. Hence the Lagrangian-maximizing contract $w(x)$ satisfies

$$w(x)=\frac{(mx-\gamma)+\lambda w_j(x)+\nu w_i(x)}{1+\lambda+\nu} \tag{8}$$

for some constant γ. Observe that increasing γ shifts the contract downwards without affecting the agent's incentive constraints, thereby increasing the principal's payoff. Since $w_i(0)=w_j(0)=0$ by assumption, it is optimal to set $\gamma=0$, which is the largest value that respects the agent's limited liability constraint.

Substituting the expression for $w(x)$ in (8) into the Lagrangian yields

$$L(\lambda,\nu):=\sup_{w\geq 0}\ \mathcal{L}(\lambda,\nu,w)=\frac{\lambda+\nu}{1+\lambda+\nu}(m\mu_i+\lambda v_{ji}+\nu v_{ii})-\lambda(v_{ii}-v_{ij}+v_{jj})-\nu v_{ii}.$$

Differentiating $L(\lambda,\nu)$ with respect to each of its arguments yields

$$\frac{dL(\lambda,\nu)}{d\lambda}=\frac{m\mu_i-v_{ji}+\nu(v_{ii}-v_{ji})}{(1+\lambda+\nu)^2}-\phi\quad\text{and}\quad\frac{dL(\lambda,\nu)}{d\nu}=\frac{m\mu_i+\lambda v_{ji}-(1+\lambda)v_{ii}}{(1+\lambda+\nu)^2}. \tag{9}$$

Although the first-order conditions need not be sufficient for a maximum (if the problem is not concave), they are necessary. We now establish the following claim.

Claim 4. *A contract solves (7) and it (strictly) dominates w_1 only if it is w_1^* or w_2^*.*

Proof of Claim 4. Observe that $dL(\lambda,\nu)/d\nu\leq 0$ if and only if $\lambda>(m\mu_i-v_{ii})/(v_{ii}-v_{ji})$, and moreover if w dominates w_1, then it must be the case that $v_{ii} - v_{ji} > 0$. To see why the last inequality is true, note that $v_{ji} \leq v_{ii} + v_{ji} - v_{ij} \leq \int w(x)dF_i(x) < v_{ii}$, where the first inequality follows from the fact that $\phi \geq 0$, the second inequality because the right-hand side of (5) is strictly positive (as argued above), and the last inequality follows from Claim 1.

It follows from the first-order conditions in (9) that one of the following pairs (λ, ν) maximizes $L(\lambda, \nu)$:

i. $\lambda = 0$ and $\nu = \infty$,

ii. $\lambda = (m\mu_i - v_{ii})/(v_{ii} - v_{ji})$ and $\nu = (v_{ii} - v_{ji})/\phi - (m\mu_i - v_{ii})/(v_{ii} - v_{ji})$, provided $(v_{ii} - v_{ji})^2 > \phi(m\mu_i - v_{ii})$, or

iii. $\lambda = \sqrt{(m\mu_i - v_{ji})/\phi} - 1$ and $\nu = 0$, provided $\sqrt{(m\mu_i - v_{ji})/\phi} - 1 > (m\mu_i - v_{ii})/(v_{ii} - v_{ji})$.

Under the first pair of multipliers, the corresponding contract is w_i, which of course cannot (strictly) payoff-dominate w_1. Recall that if w payoff-dominates w_1, then per Claim 1, it must satisfy $\int w(x)dF_i(x) < v_{ii}$. Substituting the second pair (and $\gamma = 0$) into (8) yields $\int w(x)dF_i(x) = v_{ii}$, which violates the above condition. Next, substituting the third pair of multipliers into (8) yields contract w_j^* and we have $\int w_j^*(x)dF_i(x) < v_{ii}$ (so Claim 1 may be satisfied) if and only if $\lambda > (m\mu_i - v_{ii})/(v_{ii} - v_{ji})$. Moreover, this contract (trivially) satisfies $\int w_j^*(x)dF_i(x) \geq \int w_j^*(x)dF_0(x) = 0$, which is the counterpart of $v_{3i} \geq v_{30}$ when we replace w_3 with w_j^*. This completes the proof of Claim 4. □

To conclude, we have shown that if a contract different from w_1 maximizes the principal's payoff, then this contract is either w_1^* or w_2^*. This completes the proof of Lemma 2. □

We are now ready to prove the theorem. We have shown that if there exists a contract that payoff-dominates w_1, then the payoff-maximizing contract is

$$w_j^*(x) := \rho_j w_j(x) + (1 - \rho_j)\, mx, \text{ where } \rho_j := 1 - \sqrt{\frac{\phi}{m\mu_i - v_{ji}}} \text{ for some } j \in \{1, 2\}.$$

By substituting the optimal multipliers ($\lambda = \sqrt{(m\mu_i - v_{ji})/\phi} - 1$ and $\nu = 0$) into the Lagrangian and using that strong duality holds, we have that the principal's payoff when she offers w_j^* equals[16]

$$\Pi(w_j^*) = (\sqrt{m\mu_i - v_{ji}} - \sqrt{\phi})^2. \tag{10}$$

If she offers w_1 instead, her payoff $\Pi(w_1) = m\mu_1 - v_{11}$. So $\Pi(w_j^*) > \Pi(w_1)$ only if $\sqrt{m\mu_i - v_{ji}} - \sqrt{\phi} > \sqrt{m\mu_1 - v_{11}}$ as claimed.

It remains to show that the conditions in parts (i) and (ii) of the proposition are mutually exclusive. Towards a contradiction, suppose that $\sqrt{m\mu_1 - v_{21}} - \sqrt{\phi} > \sqrt{m\mu_1 - v_{11}}$ and $\sqrt{m\mu_2 - v_{12}} - \sqrt{\phi} > \sqrt{m\mu_1 - v_{11}}\,(\geq \sqrt{m\mu_2 - v_{22}})$. We can rewrite these conditions as

$$m\mu_1 - v_{21} > m\mu_1 - v_{11} + v_{11} + v_{22} - v_{12} - v_{21} + 2\sqrt{\phi(m\mu_1 - v_{11})}\ , \text{ and}$$
$$m\mu_2 - v_{12} > m\mu_2 - v_{22} + v_{11} + v_{22} - v_{12} - v_{21} + 2\sqrt{\phi(m\mu_1 - v_{11})}\ ,$$

respectively, where we substituted $\phi = v_{11} + v_{22} - v_{12} - v_{21}$. Summing these inequalities yields $\phi + 4\sqrt{\phi(m\mu_1 - v_{11})} < 0$, which is a contradiction since both terms on the left-hand side are positive. □

A.3 Proof of Theorem 3

Consider (P′) for a given contract w_{K+1}. Fixing the dual multipliers $\lambda_{kj} \geq 0$ for each $k \in \{1, \ldots, K+1\}$ and $j \in \{0, \ldots, K+1\}$ such that $k \neq j$, the Lagrangian $L(\boldsymbol{\lambda}, w_{K+1})$

[16] Recall that whenever w_j^* payoff-dominates w_1, $m\mu_i - v_{ji} > \phi \geq 0$.

equals

$$\inf_{F_{K+1},c_{K+1},c_1,\ldots,c_K} \int (mx - w_{K+1})\, dF_{K+1} - \sum_{k=1}^{K+1}\sum_{j=0,\neq k}^{K+1} \lambda_{kj}\left(\int w_k dF_k - c_k - \int w_k dF_j + c_j\right)$$

$$= \min_x\{mx - (1+\sum_{j=0}^{K}\lambda_{K+1,j})w_{K+1}(x) + \sum_{k=1}^{K}\lambda_{k,K+1}w_k(x)\}$$

$$+ \sum_{j=0}^{K}\lambda_{K+1,j}\int w_{K+1}dF_j - \sum_{k=1}^{K}\lambda_{k,K+1}\int w_k dF_k - \sum_{k=1}^{K}\sum_{j=0,\neq k}^{K}\lambda_{kj}w_k(dF_k - dF_j)$$

$$+ \sum_{k=1}^{K+1}\begin{cases} 0 & \text{if } \lambda_{k0} + \sum_{j=1,\neq k}^{K+1}(\lambda_{kj} - \lambda_{jk}) \geq 0 \\ -\infty & \text{otherwise,} \end{cases}$$

where the last line follows from the constraint that $c_k \geq 0$ for all $k \geq 1$, and we have used that $c_0 = 0$. Notice that if $\lambda_{k0} + \sum_{j=1,\neq k}^{K+1}(\lambda_{kj} - \lambda_{jk}) < 0$ for any k, then the Lagrangian will be equal to $-\infty$, which cannot be part of an optimal solution. Therefore, we have the constraint

$$\lambda_{k0} + \sum_{j=1,\neq k}^{K+1}(\lambda_{kj} - \lambda_{jk}) \geq 0 \text{ for each } k \in \{1, \ldots, K+1\}. \tag{11}$$

By the Lagrange Duality Theorem (Luenberger, 1997, Theorem 1, p. 224), we have $\Pi(w_{K+1}) = \sup_{\boldsymbol{\lambda}\geq 0} L(\boldsymbol{\lambda}, w_{K+1})$, and so the principal's objective can be rewritten as

$$\sup\,\{L(\boldsymbol{\lambda}, w_{K+1}) \;:\; \boldsymbol{\lambda} \geq 0 \text{ and } w_{K+1} \geq 0\}.$$

Without loss, we can change the order of maximization; that is, first we maximize with respect to w_{K+1} (while holding $\boldsymbol{\lambda}$ fixed), and then we maximize with respect to $\boldsymbol{\lambda}$.

By an argument similar to that given in the proof of Theorem 2, it follows that for given multipliers $\boldsymbol{\lambda}$, the optimal contract is such that

$$w_{K+1}(x) = \frac{mx - \gamma + \sum_{k=1}^{K}\lambda_{k,K+1}w_k(x)}{1 + \sum_{j=0}^{K}\lambda_{K+1,j}}$$

for some constant γ. Notice that raising γ shifts the contract downwards, increasing the principal's payoff by γ without affecting the agent's incentives. It is thus optimal to set it to the smallest value that satisfies the agent's limited liability constraint, which is $\gamma = 0$.

Substituting w_{K+1} into the principal's objective together with (11) yields (D′). Note that the first and second constraint in (D′) corresponds to (11) for $k \in \{1, \dots, K\}$ and $k = K+1$, respectively. We have therefore shown that a contract w_{K+1} is optimal if and only if it solves (D′), and that every optimal contract takes the claimed form. □

A.4 Nature's Best Response when $K = 2$

The following corollary characterizes the additional action with which nature endows the agent when there are two known contracts and the principal optimally offers w_1^* or w_2^*; i.e., when the conditions in Theorem 2(i) or (ii) are satisfied.

Proposition 1. *Suppose that condition (i) or (ii) of Theorem 2 is satisfied so that the principal optimally offers w_j^* for some $j \in \{1, 2\}$. Then nature endows the agent with the additional action*

$$F_j^*(x) = \rho_j F_i(x) + (1 - \rho_j) F_0(x),$$

where $i \neq j$, and moreover, its cost is (weakly) smaller than $\rho_j c_i$.

That is, in response to w_1^*, nature endows the agent with an action that is a mixture of F_2 and F_0 with respective weights ρ_1 and $1 - \rho_1$ (and analogously for w_2^*). Moreover, the cost of this action is no larger than the convex combination of the actions in the mixture (so the agent prefers this action to mixing F_2 and F_0 with appropriate probability weights).

Proof of Result Proposition 1. Suppose that for some $j \in \{1, 2\}$, condition (j) of Theorem 2 is satisfied. It suffices to show that $\{w_j^*, F_j^*\}$ satisfies (5) and (6) and the principal's objective attains its maximum (which is the square of the expression given in the left-hand side of condition (j) of Theorem 2). It is straightforward to verify that $\{w_j^*, F_j^*\}$ satisfies

(5) with equality, using that $\phi = (1-\rho_j)^2(m\mu_i - v_{ji})$ by the definition of ρ_j and that $[mx - w_j(x)]dF_0(x) = 0$ by Assumption A.2.

Similarly, by substituting $\{w_j^*, F_j^*\}$ into (5), it is straightforward to show that this constraint is slack using the facts that $\int[w_j^*(x) - w_i(x)]dF_i(x) < 0$ which follows from Claim 2 in the proof of Theorem 2, and that $\int[w_j^*(x) - w_i(x)]dF_0(x) = 0$.

Next, substituting $\{w_j^*, F_j^*\}$ into the principal's objective yields

$$\int [mx - w_j^*(x)]dF_j^*(x) = \left(\sqrt{m\mu_i - v_{ji}} - \sqrt{\phi}\right)^2,$$

which is identical to the left-hand-side of condition (j) in Theorem 2.

Finally, note that incentive compatibility requires that

$$\int w_j^*(x)dF_j^*(x) - C(F_j^*) \geq 0 \text{ and } \int w_j^*(x)dF_j^*(x) - C(F_j^*) \geq \int w_j^*(x)dF_i(x) - c_i.$$

By multiplying both sides of the first constraint by $(1-\rho_j)$, both sides of the second constraint by ρ_j, and adding them, we obtain that $C(F_j^*) \leq \rho_j c_j$ as claimed. □

References

Alon, Tal, Paul Dütting, Yingkai Li, and Inbal Talgam-Cohen. 2023. "Bayesian analysis of linear contracts."

Antic, Nemanja. 2022. "Contracting with unknown technologies." *Working Paper*.

Burkett, Justin, and Maxwell Rosenthal. 2022. "Data-Driven Contract Design." *Working Paper*.

Carroll, Gabriel. 2015. "Robustness and linear contracts." *American Economic Review*, 105(2): 536–63.

Carroll, Gabriel. 2019. "Robustness in mechanism design and contracting." *Annual Review of Economics*, 11(1): 139–166.

Castro-Pires, Henrique, and Humberto Moreira. 2021. "Limited liability and non-responsiveness in agency models." *Games and Economic Behavior*, 128: 73–103.

Chade, Hector, Victoria Marone, Amanda Starc, and Jeroen Swinkels. 2022. "Multidimensional Screening and Menu Design in Health Insurance Markets." *NBER Working Paper*.

Dai, Tianjiao, and Juuso Toikka. 2022. "Robust incentives for teams." *Econometrica*, 90(4): 1583–1613.

DellaVigna, Stefano, and Devin Pope. 2018. "What motivates effort? Evidence and expert forecasts." *The Review of Economic Studies*, 85(2): 1029–1069.

Dütting, Paul, Michal Feldman, and Daniel Peretz. 2023. "Ambiguous Contracts."

Dütting, Paul, Tim Roughgarden, and Inbal Talgam-Cohen. 2019. "Simple versus optimal contracts." 369–387.

Georgiadis, George. Forthcoming. "Contracting with Moral Hazard." In *Elgar Encyclopedia on the Economics of Competition, Regulation and Antitrust.* , ed. Michael D. Noel. Cheltenham, UK: Edward Elgar Publishing.

Georgiadis, George, and Michael Powell. 2022. "A/B Contracts." *American Economic Review*, 112(1): 267–303.

Georgiadis, George, Doron Ravid, and Balazs Szentes. 2024. "Flexible Moral Hazard Problems." *Econometrica*, 92(2): 387–409.

Gottlieb, Daniel, and Humberto Moreira. 2021. "Simple contracts with adverse selection and moral hazard." *Theoretical Economics.*

Hébert, Benjamin. 2018. "Moral Hazard and the Optimality of Debt." *The Review of Economic Studies*, 85(4): 2214–2252.

Holmström, Bengt. 1979. "Moral hazard and observability." *The Bell Journal of Economics*, 74–91.

Jensen, Michael C. 2002. "Corporate budgeting is broken, let's fix it."

Lazear, Edward P. 2000. "Performance pay and productivity." *American Economic Review*, 90(5): 1346–1361.

Lazear, Edward P. 2018. "Compensation and incentives in the workplace." *Journal of Economic Perspectives*, 32(3): 195–214.

Luenberger, David G. 1997. *Optimization by vector space methods.* John Wiley & Sons.

Mirrlees, James A. 1976. "The Optimal Structure of Incentives and Authority within an Organization." *The Bell Journal of Economics*, 105–131.

Misra, Sanjog, and Harikesh S Nair. 2011. "A structural model of sales-force compensation dynamics: Estimation and field implementation." *Quantitative Marketing and Economics*, 9(3): 211–257.

Rajan, Raghuram G. 2011. "Fault Lines." In *Fault Lines*. Princeton University Press.

Rochet, Jean-Charles. 1987. "A necessary and sufficient condition for rationalizability in a quasi-linear context." *Journal of Mathematical Economics*, 16(2): 191–200.

Walton, Daniel, and Gabriel Carroll. 2022. "A General Framework for Robust Contracting Models." *Working Paper*.

Zoltners, Andris A, Prabhakant Sinha, and Sally E Lorimer. 2006. *The complete guide to sales force incentive compensation: How to design and implement plans that work.* Amacom Books.

B Additional Results (For Online Publication Only)

B.1 Optimality of linear contracts when the agent is known to earn sufficiently large rents under a single known contract

The following result shows that with one known contract (i.e., $K = 1$), if the agent is known to earn sufficiently large rents, then consistent with most of the robust contracting literature (e.g., (Carroll, 2015)), a linear contract is optimal.

Result 1. *Assume there is one known contract, w_1, and the principal knows that c_1 is no larger than $\widehat{c}$. If $\widehat{c}$ is smaller than some threshold $\bar{c}$, then the linear contract with slope $\sqrt{m\widehat{c} \,/ \int x dF_1(x)}$ maximizes the principal's worst-case profit. Otherwise w_1 is optimal.*

Unlike in the argument for Theorem 1, because the cost of F_1 is bounded, the principal can take away some of the agent's rents while dissuading him from switching to F_0. Of course, there are other actions that nature could endow the agent with. Linear contracts ensure that actions which are appealing to the agent are not very harmful to the principal per the standard intuition; see, for example, Carroll (2015).

When the principal knows that the agent is getting a significant utility from w_1 and action F_1, some of that utility can be taken away and the agent will not switch to taking action F_0. By offering a linear contract, the principal is also protecting herself from other actions that nature could endow the agent, since any action benefiting the agent will also benefit the principal.

Proof of Result 1. Assume there is one known contract, and moreover, the principal knows that the cost of F_1 is no larger than $\widehat{c} \leq \int w_1(x) dF_1(x)$. Her problem can be expressed as follows:

$$\begin{aligned}\sup_{w_2\geq 0}\ \inf_{F_2,c_1,c_2}\ & \int [mx - w_2(x)]dF_2(x) \\ \text{s.t.}\ & \int w_2(x)dF_2(x) - c_2 \geq \int w_2(x)dF_1(x) - c_1 \\ & \int w_1(x)dF_1(x) - c_1 \geq \int w_1(x)dF_2(x) - c_2 \\ & \int w_2(x)dF_2(x) - c_2 \geq \int w_2(x)dF_0(x) \\ & c_1 \leq \widehat{c},\ c_2 \geq 0, \text{ and } F_2 \in \Delta(\mathcal{X})\end{aligned} \tag{12}$$

Since $\widehat{c} \leq \int w_1 dF_1$, the agent prefers F_1 to F_0 when w_1 is offered. We guess (and later verify) that $w_2(0) = 0$, so that the third constraint becomes $\int w_2 dF_2 - c_2 \geq 0$.

Fix some w_2. We have the Lagrangian

$$\begin{aligned}L(\boldsymbol{\lambda}, w_2) = \inf & \int [mx - (1+\lambda_1+\lambda_3)w_2(x) + \lambda_2 w_1(x)]\, dF_2(x) \\ & + \int [\lambda_1 w_2 - \lambda_2 w_1]\, dF_1 - (\lambda_1-\lambda_2)c_1 - (\lambda_2-\lambda_1-\lambda_3)c_2 \\ & \text{s.t. } F_2 \in \Delta(\mathcal{X}),\ 0 \leq c_1 \leq \widehat{c} \text{ and } \quad c_2 \geq 0 \\ & = \min\{mx - (1+\lambda_1+\lambda_3)w_2(x) + \lambda_2 w_1(x)\} \\ & + \int (\lambda_1 w_2 - \lambda_2 w_1)dF_1 - [\lambda_1-\lambda_2]^+\,\widehat{c} - \begin{cases} 0 & \text{if } \lambda_2 \leq \lambda_1+\lambda_3 \\ \infty & \text{if } \lambda_2 > \lambda_1+\lambda_3, \end{cases}\end{aligned}$$

where $\lambda_1, \lambda_2, \lambda_3 \geq 0$ are the dual multipliers corresponding to the first, second, third constraint of (12), respectively. To derive the fourth line we used that the integral in the first line is maximized by the degenerate distribution, F_2, which places all mass at $\hat{x} \in \min_x \{mx - (1+\lambda_1+\lambda_3)w_2(x) + \lambda_2 w_1(x)\}$.

Notice that holding w_2 fixed, nature's problem is linear. Hence by the Lagrange Duality Theorem (Luenberger, 1997, Theorem 1, p. 224) strong duality holds, and so the solution to (12) equals

$$\sup_{w_2 \geq 0, \boldsymbol{\lambda} \geq 0} \min \{mx - (1 + \lambda_1 + \lambda_3) w_2(x) + \lambda_2 w_1(x)\} \tag{13}$$
$$+ \int [\lambda_1 w_2(x) - \lambda_2 w_1(x)]\, dF_1(x) - [\lambda_1 - \lambda_2]^+ \widehat{c}$$
$$\text{s.t. } \lambda_2 \leq \lambda_1 + \lambda_3,$$

where the constraint follows from the observation that if $\lambda_2 \nleq \lambda_1 + \lambda_3$, then the objective equals $-\infty$, which cannot be optimal. We shall maximize (13) first with respect to w_2 and then with respect to $\boldsymbol{\lambda}$. Observe that for any given $\boldsymbol{\lambda}$, the optimal contract must be such that the term inside the curly brackets in (13) is constant in x, that is,

$$w_2(x) = \frac{mx + \gamma + \lambda_2 w_1(x)}{1 + \lambda_1 + \lambda_3}$$

for some $\gamma \geq 0$. Since this constant shifts the contract by the same amount for all x, it does not affect the agent's incentives, and so it is optimal to set $\Gamma = 0$. Using this expression we can rewrite (13) as

$$\sup_{\boldsymbol{\lambda} \geq 0} \frac{\lambda_1 m \mu_1 - (1 + \lambda_3) \lambda_2 v_{11}}{1 + \lambda_1 + \lambda_3} - [\lambda_1 - \lambda_2]^+ \widehat{c} \quad \text{s.t.} \quad \lambda_1 + \lambda_3 \geq \lambda_2 \geq 0,$$

where $\mu_1 := \int x dF_1(x)$ and $v_{11} := \int w_1 dF_1$.

For any fixed $\lambda_1, \lambda_3 \geq 0$, the objective increases in λ_2 at rate

$$-\frac{1 + \lambda_3}{1 + \lambda_1 + \lambda_3} v_{11} + \widehat{c} \text{ if } \lambda_2 \leq \lambda_1, \text{ and}$$
$$-\frac{1 + \lambda_3}{1 + \lambda_1 + \lambda_3} v_{11} < 0 \text{ if } \lambda_2 > \lambda_1,$$

which implies that the objective is maximized by setting

$$\lambda_2 = \begin{cases} 0 & \text{if } (1+\lambda_1+\lambda_3)\widehat{c} < (1+\lambda_3)v_{11}, \text{ and} \\ \lambda_1 & \text{if } (1+\lambda_1+\lambda_3)\widehat{c} > (1+\lambda_3)v_{11}. \end{cases} \tag{14}$$

Because $\lambda_2 \leq \lambda_1$, the constraint in (13) is satisfied for all $\lambda_3 \geq 0$, and since the objective decreases in λ_3, it is optimal to set $\lambda_3 = 0$.

Using (14) and $\lambda_3 = 0$, we can write the objective solely as a function of λ_1 as

$$\sup_{\lambda_1 \geq 0} \frac{\lambda_1 m\mu_1}{1+\lambda_1} - \lambda_1\widehat{c} + \lambda_1\left[\widehat{c} - \frac{v_{11}}{1+\lambda_1}\right]^+.$$

This objective increases in λ_1 at rate

$$\begin{aligned} m\mu_1/(1+\lambda_1)^2 - \widehat{c} &\quad \text{for } \lambda_1 < v_{11}/\widehat{c} - 1, \text{ and} \\ (m\mu_1 - v_{11})/(1+\lambda_1)^2 > 0 &\quad \text{for } \lambda_1 \geq v_{11}/\widehat{c} - 1. \end{aligned}$$

That is, the objective is initially concave, peaking at $\sqrt{m\mu_1/\widehat{c}} - 1$ (provided this is smaller than $v_{11}/\widehat{c} - 1$), it has a kink at $\lambda_1 = v_{11}/\widehat{c} - 1$, and is then increasing. We thus have two candidates for the optimal value of λ_1: $\sqrt{m\mu_1/\widehat{c}} - 1$ and ∞. The principal's objective evaluated at the first and second candidate is $(\sqrt{m\mu_1} - \sqrt{\widehat{c}})^2$ and $(m\mu_1 - v_{11})$, respectively, and comparing the two yields that the optimal

$$\lambda_1 = \begin{cases} \sqrt{m\mu_1/\widehat{c}} - 1 & \text{if } \widehat{c} \leq \bar{c} := 2m\mu_1 - \sqrt{2m\mu_1 - v_{11}}, \text{ and} \\ \infty & \text{otherwise.} \end{cases}$$

In the former case, $\lambda_2 = 0$ and so $w_2(x) \equiv \sqrt{m\widehat{c}/\mu_1}\, x$ is optimal. In the latter case, $\lambda_2 = \lambda_1$ and so $w_2(x) \equiv w_1(x)$ is optimal. $\square$