Robustness and Linear Contracts

Gabriel Carroll, Microsoft Research and Stanford University
gabriel.d.carroll@gmail.com

December 21, 2012

Abstract

We consider a simple moral hazard problem, under risk-neutrality and limited liability, in which the principal is uncertain about the technology available to the agent. The principal knows some actions available to the agent, but other, unknown actions may also exist. The principal evaluates contracts according to their worst-case performance, with respect to the actions that may or may not be available to the agent. Under very general circumstances, the unique optimal contract is linear. This model thus provides a new explanation for the widespread use of linear contracts in practice, as well as a flexible and tractable modeling approach for moral hazard under non-quantifiable uncertainty.

Thanks to (in random order) Daron Acemoglu, Luis Zermeño, Iván Werning, Alex Wolitzky, Mike Riordan, Lucas Maestri, Ben Golub, Sylvain Chassang, Abhijit Banerjee, Bengt Holmström, and Tomasz Sadzik for helpful discussions and advice, as well as other seminar participants at Zurich, UCLA, Toulouse, Caltech, and Harvard EconCS for their comments.

1 Introduction

This paper considers a simple principal-agent problem with uncertain technology. As in the usual framework, the agent takes an unobserved costly action, which stochastically determines output. The agent can be paid based on observed output. The principal wishes

to maximize the expectation of output minus the wage paid to the agent. Where this paper deviates from most of the principal-agent literature is in the principal's knowledge of the set of actions available to the agent (including their associated costs). Rather than assuming that this set of actions is known, as is common, we assume that the principal knows some available actions, but other, unknown actions may also exist. The principal does not have a prior belief about which actions exist. Instead, she evaluates possible contracts using a worst-case criterion: any contract is evaluated by its worst performance over all sets of actions consistent with her knowledge. The principal and agent are financially risk-neutral, and payments are constrained by limited liability. We show that, under broad conditions, the unique optimal contract is linear: the agent is paid a fixed share of output.

The importance of this finding can be viewed in three different ways. First, it addresses a longstanding problem in contract theory: why are linear contracts so common? The model here offers a simple and general new explanation. It answers the call made by Holmström and Milgrom in their classic paper on linear contracts in dynamic environments [12, p. 326]:

> It is probably the great robustness of linear rules based on aggregates that accounts for their popularity. That point is not made as effectively as we would like by our model; we suspect that it cannot be made effectively in any traditional Bayesian model. But issues of robustness lie at the heart of explaining any incentive scheme which is expected to work well in practical environments.

The second view of our contribution is that it provides concrete advice to people faced with the practical task of designing incentive contracts under non-quantifiable uncertainty. And, third, it offers a flexible analytical framework that can be used to model more complex moral hazard problems in a tractable way.

Mathematically, the main result of this paper is rather simple. What is surprising is that it did not appear much earlier in the agency theory literature. Accordingly, the paper aims to fill a longstanding methodological gap in this literature, presenting a simple formal model for studying robustness in incentive contracts that may help in understanding more complex agency issues.

The intuition for the main result is as follows. When the principal proposes a contract, in the face of her uncertainty about the agent's technology, she knows very little about what will happen; but the one thing she does know is a lower bound on the agent's

expected payoff (due to the actions that are known to be available). The only useful way to turn this into a lower bound on her own expected payoff is to impose a linear relationship between the two payoffs. Some related intuitions have appeared in the previous literature on linear contracts (discussed in more detail in the concluding section), but the idea seems to have never been presented in the form here.

Section 2 of the paper formally presents the basic version of the model and result. The model is kept simple here, to illustrate the crucial ideas as cleanly as possible. Section 3 shows how the logic of the result persists under various extensions that enrich the model or make it more realistic. This includes allowing a participation constraint, incorporating risk-aversion, and assuming some known limits on the actions available to the agent, as well as allowing the principal to screen agents on their technology.

Part of the message of this paper is to illustrate how worst-case objectives can provide a tractable alternative to fully Bayesian objectives in mechanism design. Accordingly, this paper joins a growing literature on maxmin mechanism design. This includes the work of Hurwicz and Shapiro [13], Frankel [9], and Garrett [10], also on contracting with unknown agent preferences; the work initiated by Bergemann and Morris [1] and Chung and Ely [6] on mechanism design with unknown higher-order beliefs; and work such as Yamashita's [19] on maxmin expected welfare under weak assumptions on agent behavior (assuming only that agents play undominated strategies). The implementation literature (surveyed in [15]) can also be seen as mechanism design with a worst-case rather than Bayesian objective: it seeks to construct mechanisms that ensure a desirable outcome in *all* equilibria. A broader mechanism design literature provides nearly optimal worst-case performance in various settings, without optimizing exactly; recent examples include [4, 18, 17].

This paper also adds to the literature on foundations of linear contracts, and some discussion of its relationship to that literature is in order. However, in order to come to the model more quickly, we defer that discussion to the concluding section. In any case, the literature can be discussed more concretely after presenting the model in detail.

2 The basic model

2.1 Notation

We write $\Delta(X)$ for the space of Borel distributions on $X \subseteq \mathbb{R}^k$, equipped with the weak topology. For $x \in X$, δ_x is the degenerate distribution putting probability 1 on x. $\mathbb{R}^+$ is

the set of nonnegative reals.

2.2 Setup

A principal contracts with an agent, who is to take a costly action that produces a stochastic output. The action is not observable to the principal; only the resulting output, y, is observable. Thus, payment to the agent can depend only on y, and this dependence is what provides incentives. Both parties are financially risk-neutral.

We write Y for the set of possible output values, and assume Y is a compact subset of $\mathbb{R}$. Y may be finite or infinite. We normalize $\min(Y) = 0$.

An *action* is represented by a pair $(F, c) \in \Delta(Y) \times \mathbb{R}^+$. The interpretation is that the action leads to distribution F over output, and costs c to the agent. c may be interpreted literally as a monetary cost, or an additive disutility of effort; we take no stand on this. We give $\Delta(Y) \times \mathbb{R}^+$ the natural product topology. A *technology* is a compact subset of $\Delta(Y) \times \mathbb{R}^+$. The technology describes the set of actions available to the agent. The agent knows his technology $\mathcal{A}$, but the principal does not. Instead, the principal knows only some set $\mathcal{A}_0$ of actions available to the agent, and she believes A may be any (compact) superset of $\mathcal{A}_0$.

The exogenous $\mathcal{A}_0$ may be any technology, subject to the following *nontriviality* assumption: There exists $(F, c) \in \mathcal{A}_0$ such that $E_F[y] - c > 0$. This assumption ensures that the principal benefits from hiring the agent.

It is natural to assume that the agent can always exert no effort; this corresponds to assuming $(\delta_0, 0) \in \mathcal{A}_0$. However our results will not formally require this assumption. Also, we say $\mathcal{A}_0$ satisfies the *full-support condition* if, for all $(F, c) \in \mathcal{A}_0$ such that $(F, c) \neq (\delta_0, 0)$, F has full support on Y. This assumption will strengthen our main result.

On to contracts, which specify how much the agent is paid for each level of output. We assume one-sided limited liability: the agent can never be paid less than zero. Thus, a *contract* is any continuous function $w : Y \to \mathbb{R}^+$.[1]

We can now summarize the timing of the game:

1. the principal offers a contract w;
2. the agent, knowing $\mathcal{A}$, chooses action $(F, c) \in \mathcal{A}$;
3. output $y \sim F$ is realized;

[1] Requiring continuity ensures the agent's optimization problem has a solution. If, say, Y is an arbitrarily fine discrete grid, then continuity is a vacuous assumption.

4. payoffs are received: $y - w(y)$ to the principal and $w(y) - c$ to the agent.

Describing the agent's behavior is simple, since he maximizes expected utility. Given contract w, and technology $\mathcal{A}$, the agent's choice set is

$$A^*(w|\mathcal{A}) = \arg \max_{(F,c)\in\mathcal{A}} (E_F[w(y)] - c)\,.$$

Continuity and compactness ensure this set is nonempty. It will also be useful to write

$$V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}} (E_F[w(y)] - c)$$

for the agent's expected payoff. If the agent is indifferent among several actions, we assume he maximizes the principal's utility (as is common in this literature). Thus the principal's expected payoff under technology $\mathcal{A}$ is

$$V_P(w|\mathcal{A}) = \max_{(F,c)\in A^*(w|\mathcal{A})} E_F[y - w(y)].$$

Finally, we assume the principal evaluates contracts by their worst-case expected payoff, over all possible technologies $\mathcal{A}$:

$$V_P(w) = \inf_{\mathcal{A}\supseteq\mathcal{A}_0} V_P(w|\mathcal{A}).$$

Our focus is on the principal's problem, namely to maximize $V_P(w)$. In the next section, we will show that the maximum exists, and identify the contract that attains it.

2.3 Analysis

In the above model, the principal considers the worst case over a very wide range of technologies. Faced with this huge uncertainty, can she even guarantee herself a positive expected payoff? Yes; in fact *linear* contracts — those of the form $w(y) = \alpha y$ for constant α — can provide this guarantee. To see this, suppose the principal offers such a contract, with $\alpha > 0$. (We can also assume $\alpha \leq 1$, since clearly $\alpha > 1$ cannot earn a positive payoff.) Note that whatever technology $\mathcal{A} \supseteq \mathcal{A}_0$ the agent has, and whatever optimal action (F, c) he chooses,

$$\alpha \cdot E_F[y] \geq E_F[w(y)] - c = V_A(w|\mathcal{A}) \geq V_A(w|\mathcal{A}_0),$$

since his action is optimal in $\mathcal{A}$, a superset of $\mathcal{A}_0$. And so the principal's payoff is

$$E_F[y - w(y)] = (1 - \alpha) \cdot E_F[y] \geq \frac{1-\alpha}{\alpha} V_A(w|\mathcal{A}_0).$$

Since this holds regardless of the technology,

$$V_P(w) \geq \frac{1-\alpha}{\alpha} V_A(w|\mathcal{A}_0) = \max_{(F,c)\in\mathcal{A}_0} \left((1-\alpha)E_F[y] - \frac{1-\alpha}{\alpha} c \right). \tag{2.1}$$

The nontriviality assumption implies that if α is close to 1 then $V_A(w|\mathcal{A}_0) > 0$, and so we have a positive lower bound on the principal's worst-case payoff.

This shows how to obtain a payoff guarantee from a linear contract. But is it possible that some other, subtler contract form would give a better guarantee? The answer is no, and we give now a sketch of the argument.

Consider any arbitrary contract w. It implies some guaranteed (expected) payoff to the agent regardless of his technology, namely $V_A(w|\mathcal{A}_0)$, and some guaranteed payoff to the principal, namely $V_P(w)$. For the sake of concrete illustration, we arbitrarily choose numbers; say for example that the agent's guarantee is 123 and the principal's is 456. Given the uncertainty about the technology, from the principal's point of view, the agent may potentially take any action — with one constraint: the expected payment must be at least 123, since she knows for sure he can earn at least this much. Thus, the fact that the principal's guarantee is 456 means that any distribution F over outcomes satisfying $E_F[w(y)] \geq 123$ must also satisfy $E_F[y - w(y)] \geq 456$. Applying a separation theorem, we conclude that there exist constants κ, λ such that $y - w(y) \geq \kappa w(y) + \lambda$ for all y, and $456 \leq \kappa \cdot 123 + \lambda$. In particular, the guarantee of contract w (for the principal) derives entirely from these linear inequalities.

Now we construct w' such that $y - w'(y) \geq \kappa w'(y) + \lambda$, and check that $w'(y) \geq w(y)$. This implies that w' also guarantees at least 123 for the agent, and so at least $\kappa \cdot 123 + \lambda \geq 456$ for the principal. Moreover w' is a linear contract (or more precisely, an affine contract). This argument shows that for any contract, there is a linear contract that performs at least as well.

We now proceed to fill in the details of the argument. But we first remark in passing that there is also another fast way to show that any contract is (weakly) outperformed by a linear contract; this alternative proof, with some discussion, is in Appendix A.

The first step of the argument is to exactly identify the guarantee $V_P(w)$ from any given contract w. The characterization (Lemma 2.1 below) is intuitive: find the worst

conceivable output distribution for the principal, subject to only one constraint, namely the known lower bound on the agent's payoff. However, the full proof is slightly more involved because the assumption of tie-breaking in favor of the principal introduces some technicalities.

One such issue is that we must deal with the zero contract ($w(y) = 0$ for all y) separately. We abusively denote this contract by 0. Suppose there exists $(F, c) \in \mathcal{A}_0$ with $c = 0$; that is, the agent can definitely produce some output costlessly. Then the principal's guarantee is simply the highest value of $E_F[y]$ over such F. Formally, $A^*(0|\mathcal{A}) = \{(F, c) \in \mathcal{A} \mid c = 0\}$; then $V_P(0|\mathcal{A}) = \max_{(F,0)\in\mathcal{A}} E_F[y]$, and so $V_P(0) = \max_{(F,0)\in\mathcal{A}_0} E_F[y]$. If there is no action $(F, 0) \in \mathcal{A}_0$, then the principal is not guaranteed any positive payoff: taking $\mathcal{A} = \mathcal{A}_0 \cup \{(\delta_0, 0)\}$, we have $A^*(0|\mathcal{A}) = \{(\delta_0, 0)\}$, hence $V_P(0) = 0$.

Now we can focus on contracts that perform better than the zero contract.

Lemma 2.1. *Let w be any nonzero contract such that $V_P(w) \geq V_P(0)$. Then,*

$$V_P(w) = \min E_F[y - w(y)] \quad \text{over } F \in \Delta(Y) \text{ such that } E_F[w(y)] \geq V_A(w|\mathcal{A}_0). \tag{2.2}$$

Moreover, as long as $V_P(w) > 0$, then for any F attaining the minimum, the condition holds with equality: $E_F[w(y)] = V_A(w|\mathcal{A}_0)$.

We include the proof here for completeness, but it can be skipped on a first reading.

Proof: First, consider any technology $\mathcal{A} \supseteq \mathcal{A}_0$. The agent's payoff is at least $V_A(w|\mathcal{A}_0)$. That is, he chooses an action (F, c) such that

$$E_F[w(y)] \geq E_F[w(y)] - c \geq V_A(w|\mathcal{A}_0).$$

Hence the principal's payoff, $V_P(w|\mathcal{A}) = E_F[y - w(y)]$, is at least the minimum given by (2.2). Thus, the principal's worst-case payoff $V_P(w)$ is no lower than given by (2.2).

To see this is tight, let F be a distribution attaining the minimum in (2.2). First suppose that F does not place full support on values of y for which w attains its maximum. Then let F' be a mixture of F with weight $1-\epsilon$, and a mass point δ_{y^*} with weight ϵ, where y^* is some point where w attains its maximum. Then $E_{F'}[w(y)] > E_F[w(y)] \geq V_A(w|\mathcal{A}_0)$. The strict inequality means that if $\mathcal{A} = \mathcal{A}_0 \cup \{(F', 0)\}$, then the agent's unique optimal action in $\mathcal{A}$ is $(F', 0)$, leading to expected payoff $(1 - \epsilon)E_F[y - w(y)] + \epsilon(y^* - w(y^*))$ for the principal. As $\epsilon \to 0$ this converges to the minimum in (2.2), so the principal cannot be guaranteed any higher expected payoff.

Now suppose F does place full support on values of y at which w attains its maximum. If $E_F[w(y)] > V_A(w|\mathcal{A}_0)$, then we can again proceed as above with $\mathcal{A} = \mathcal{A}_0 \cup \{(F, 0)\}$. This leaves only the case of equality — $V_A(w|\mathcal{A}_0) = \max_y w(y)$ — which is only satisfied when $\mathcal{A}_0$ contains some action of the form $(F, 0)$ with F supported at output levels for which w attains its maximum, and under technology $\mathcal{A}_0$ the agent must choose such an action. Then the principal's expected payoff $V_P(w|\mathcal{A}_0)$ is the maximum of $E_F[y] - \max_y w(y)$ over all such actions $(F, 0)$. But this is less than $E_F[y] \leq \max_{(F,0)\in\mathcal{A}_0} E_F[y] = V_P(0)$. Thus we have $V_P(w) < V_P(0)$, contradicting the given assumption.

This shows (2.2). Now assume $V_P(w) > 0$, and let $F \in \Delta(Y)$ attain the minimum in (2.2). We have $E_F[y - w(y)] = V_P(w) > 0$. On the other hand, $y - w(y) \leq 0$ when $y = 0$. Now if we have $E_F[w(y)] > V_A(w|\mathcal{A}_0)$ strictly, then replace F by a mixture of F with weight $1 - \epsilon$ and δ_0 with weight ϵ, for small ϵ, to see that minimality is contradicted. Hence we have equality, $E_F[w(y)] = V_A(w|\mathcal{A}_0)$, as claimed. □

Note that the equality statement in Lemma 2.1 implies that (2.1), the guarantee of a linear contract, is actually an equality. We record this as a separate lemma:

Lemma 2.2. *For any $\alpha > 0$, if the guarantee of the linear contract $w(y) = \alpha y$ satisfes $V_P(w) \geq V_P(0)$ and $V_P(w) > 0$, then*

$$V_P(w) = \max_{(F,c)\in\mathcal{A}_0} \left((1 - \alpha)E_F[y] - \frac{1 - \alpha}{\alpha} c \right). \tag{2.3}$$

This remains valid for $\alpha = 0$, if we interpret the second term as 0 for $c = 0$ and $-\infty$ for $c > 0$.

Now we are ready for the main result — the optimality of linear contracts.

Theorem 2.3. *There exists a linear contract w that maximizes V_P. Moreover, if $\mathcal{A}_0$ satisfies the full-support condition, then* every *contract that maximizes V_P is linear.*

The proof follows the sketch given earlier: for any proposed contract w, we use a separation argument to find a linear inequality that underlies the principal's guarantee; we then find a linear contract w' that satisfies the same inequality and is more generous to the agent, which means it must guarantee more to the principal as well.

Proof: Let w be any contract that does weakly better than the zero contract and has strictly positive guarantee: $V_P(w) \geq V_P(0)$ and $V_P(w) > 0$. As sketched above, we will show that there is a linear contract w' that does weakly better than w, and that under the full-support condition, w' does strictly better unless $w' = w$. We may assume that w is not linear (otherwise take $w' = w$).

Let $S \subseteq \mathbb{R}^2$ be the convex hull of all points $(w(y), y - w(y))$ for $y \in Y$. Let T be the set of all pairs $(u, v) \in \mathbb{R}^2$ such that $u > V_A(w|\mathcal{A}_0)$ and $v < V_P(w)$. The conclusion (2.2) of Lemma 2.1 implies that S and T are disjoint. So by the separating hyperplane theorem, there exist constants λ, μ, ν such that

$$\lambda u + \mu v \leq \nu \qquad \text{for all} \qquad (u, v) \in S, \tag{2.4}$$

$$\lambda u + \mu v \geq \nu \qquad \text{for all} \qquad (u, v) \in T, \tag{2.5}$$

and $(\lambda, \mu) \neq (0, 0)$. In addition, if we let F^* be the distribution attaining the minimum in (2.2), the pair $(E_{F^*}[w(y)], E_{F^*}[y - w(y)])$ lies in the closures of both S and T, hence

$$\lambda E_{F^*}[w(y)] + \mu E_{F^*}[y - w(y)] = \nu. \tag{2.6}$$

Condition (2.5) implies $\lambda \geq 0$ and $\mu \leq 0$. Let us show that both these inequalities hold strictly. If $\mu = 0$, then $\lambda > 0$, and (2.4) and (2.5) imply $\max_{y \in Y} w(y) \leq \nu/\lambda \leq V_A(w|\mathcal{A}_0)$. But then as in the proof of Lemma 2.1 (using the fact that w is not the zero contract, since w is assumed nonlinear) we obtain $V_P(w) < V_P(0)$, a contradiction. If $\lambda = 0$, then $\mu < 0$, and (2.4) and (2.5) imply $\min_{y \in Y}(y - w(y)) \geq \nu/\mu \geq V_P(w)$. But $\min_{y \in Y}(y - w(y)) \leq 0 - w(0) \leq 0$, so $V_P(w) \leq 0$, again contrary to assumption.

Now, inequality (2.4), applied to each pair $(w(y), y - w(y))$, can be rearranged as

$$w(y) \leq \frac{\nu - \mu y}{\lambda - \mu}.$$

Now define

$$w'(y) = \frac{\nu - \mu y}{\lambda - \mu}.$$

Thus $w' \geq w$ pointwise. Notice that this immediately implies $w'(y) \geq 0$ for all y. So w' is indeed a contract.

We will show that $V_P(w') \geq V_P(w)$. For any action (F, c) taken by the agent under contract w' and any technology $\mathcal{A}$, we have

$$E_F[w'(y)] \geq E_F[w'(y)] - c = V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0). \tag{2.7}$$

Using the linear relation

$$y - w'(y) = \frac{\lambda w'(y) - \nu}{-\mu}$$

for each y, we obtain

$$V_P(w'|\mathcal{A}) = E_F[y - w'(y)] \geq \frac{\lambda V_A(w'|\mathcal{A}_0) - \nu}{-\mu}. \tag{2.8}$$

On the other hand, rearranging (2.6) and using the last statement of Lemma 2.1,

$$\frac{\lambda V_A(w|\mathcal{A}_0) - \nu}{-\mu} = \frac{\lambda E_{F^*}[w(y)] - \nu}{-\mu} = E_{F^*}[y - w(y)] = V_P(w). \tag{2.9}$$

Combining (2.8) and (2.9) gives

$$V_P(w'|\mathcal{A}) \geq V_P(w) + \frac{\lambda}{-\mu}\left(V_A(w'|\mathcal{A}_0) - V_A(w|\mathcal{A}_0)\right) \geq V_P(w). \tag{2.10}$$

Since this holds for all $\mathcal{A}$, then, $V_P(w') \geq V_P(w)$.

If the full-support condition holds, then let (F, c) be the action taken under w and technology $\mathcal{A}_0$. We have $(F, c) \neq (\delta_0, 0)$ (otherwise $V_P(w) \leq 0$), so F has full support. Therefore, if w and w' do not coincide, we have $E_F[w'(y)] > E_F[w(y)]$ implying $V_A(w'|\mathcal{A}_0) > V_A(w|\mathcal{A}_0)$. Then (2.10) shows that $V_P(w') > V_P(w)$ strictly.

The above shows that, given w, there is an affine contract w' — that is, one of the form $w'(y) = \alpha y + \beta$ — that does weakly better than w, and strictly better if the full-support condition holds and $w' \neq w$; and that satisfies $w' \geq w$ pointwise. In particular $\beta = w'(0) \geq 0$. Now replace $w'(y)$ by $\alpha y = w'(y) - \beta$; this further increases $V_P(w')$ by β, since a constant shift does not affect the agent's incentives for choice of action. Thus we get a linear contract that does weakly better than w, and strictly better if the full-support condition holds and w was not already linear.

Now we are basically done. Our first assertion to prove was that there exists an optimal contract that it is linear. We first check that within the class of linear contracts, there is an optimal one. Recall the formula (2.3); this expression is continuous in the share $\alpha \in [0, 1]$, so it achieves a maximum. We saw in (2.1) that the guarantee of the linear contract with share α is at least the formula in (2.3), and since equality holds whenever the contract guarantees at least as much as the zero contract and a positive amount, we see that the maximum of (2.3) is in fact the optimal guarantee among linear contracts. Now, the preceding argument shows that no nonlinear contract can do better than all linear contracts, so the optimal linear contract is in fact optimal among all possible contracts.

Furthermore, suppose the full-support condition holds. If there is a nonlinear contract w that is also optimal, then the above argument shows that there is some linear contract

that strictly outperforms w, a contradiction. $\square$

To complete the analysis, we may as well explicitly identify the share α in the optimal linear contract. From Lemma 2.2, the optimal share is found by maximizing

$$(1-\alpha)E_F[y] - \frac{1-\alpha}{\alpha}c$$

jointly over $(F,c) \in \mathcal{A}_0$ and $\alpha \in [0,1]$. When $E_F[y] < c$, the maximum value is 0 (given by $\alpha = 1$). Otherwise, maximizing over α gives $\alpha = \sqrt{c/E_F[y]}$, and the objective reduces to

$$E_F[y] + c - 2\sqrt{cE_F[y]} = (\sqrt{E_F[y]} - \sqrt{c})^2. \tag{2.11}$$

Thus, the optimal contract is chosen by taking $(F^*, c^*) \in \mathcal{A}_0$ to maximize this expression, subject to $E_F[y] \geq c$, and then choosing $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$ to be the share. If it happens that there are several actions in $\mathcal{A}_0$ attaining the maximum (a knife-edge case), then there are several optimal linear contracts.

Finally, we comment on the role of some assumptions in the model. The uncertainty on the principal's part is clearly essential: If the principal knew for certain that $\mathcal{A} = \mathcal{A}_0$, then the optimal contract would in general not be linear (see e.g. [7]). For example, if Y is finite and $\mathcal{A}$ contains only two actions, the optimal way to incentivize the costlier action is to pay a positive amount only for the value of output whose likelihood ratio is highest.

The limited liability assumption is also crucial. If we removed this assumption, and instead constrained payments from below by imposing a participation constraint (say, the agent must be assured a nonnegative expected payoff), then the standard solution of "selling the firm to the agent" would apply: clearly the principal could not be guaranteed any higher payoff than the total surplus under $\mathcal{A}_0$, $s_0 = \max_{(F,c)\in\mathcal{A}_0}(E_F[y] - c)$, and could achieve this payoff by setting $w(y) = y - s_0$.

3 Extensions

In this section we consider several variations of the basic model. The purpose is twofold: to study how the result persists when the model is made more realistic, and to show how the analytical tools extend to more complex models.

Specifically, we consider adding a participation constraint, reducing the range of possible technologies by limiting the possible actions or assuming the principal knows a lower

bound on the cost of any given distribution, and including risk aversion. Finally, we also consider a version where the principal can screen by offering different contracts depending on the agent's technology $\mathcal{A}$, rather than offering just one contract.

3.1 Participation constraint

In the basic model, only the limited liability constraint disciplined payments from below. We could imagine that there is also a participation constraint, so that the principal is required to guarantee the agent an expected payoff of at least $\overline{U}_A > 0$. This could be incorporated by restricting the principal's maximization problem to contracts w satisfying $E_F[w(y)] - c \geq \overline{U}_A$ for some $(F, c) \in \mathcal{A}_0$. Let us assume there exists such a w satisfying $V_P(w) \geq V_P(0)$ and $V_P(w) > 0$.

In this case, the same argument as before would show that there exists an optimal contract that is affine — that is, $w(y) = \alpha y + \beta$ for some constants α, β with $\alpha \in [0, 1]$ — and that under the full-support condition every optimal contract is affine. Indeed, since the contract w' constructed in the proof of Theorem 2.3 satisfies $w' \geq w$ everywhere, if w satisfies the participation constraint, so does w'. The only step of the original proof that does not go through is changing from $w'(y) = \alpha y + \beta$ to αy, since the latter contract may not satisfy the constraint.

However, we can do better. For any given α, the optimal choice of β is to be as small as possible subject to the nonnegativity and participation constraints:

$$\beta = \max\{0, \overline{U}_A - \max_{(F,c)\in\mathcal{A}_0} (\alpha E_F[y] - c)\}. \tag{3.1}$$

Let us show that a choice of α such that $\max_{(F,c)}(\alpha E_F[y] - c) < \overline{U}_A$ — so that $\beta > 0$ — cannot be optimal. If it is optimal, the principal's guarantee equals the guarantee from Lemma 2.2 minus β, which simplifies to

$$\max_{(F,c)\in\mathcal{A}_0} \left(E_F[y] - \frac{1}{\alpha} c \right) - \overline{U}_A. \tag{3.2}$$

We must have $\alpha < 1$ (since $V_P(w) > 0$ at the optimum by assumption). If we increase α slightly, and adjust β so that (3.1) still holds, the principal's guarantee (3.2) increases, contradicting optimality. Intuitively, as long as the payments are being constrained below by the participation constraint rather than by limited liability, the principal would like to align the agent's incentives with her own interests as much as possible.

This shows that the optimal contracts must be linear, $w(y) = \alpha y$, just as before.

3.2 Alternative sets of technologies

The model as written specifies that the principal considers any technology $\mathcal{A} \supseteq \mathcal{A}_0$ to be possible. Such drastic uncertainty regarding the technology might be unrealistic. However, all of the same results hold if the principal considers a much smaller set of possible technologies $\mathcal{A}$: either $\mathcal{A}_0$ itself, or $\mathcal{A}_0$ with just one more action (F, c) added. To see this, just check that when $V_P(w)$ is redefined as the infimum of $V_P(w|\mathcal{A})$ over this restricted set of technologies, its value does not change.

In fact, we do not even need to assume that there is a single minimal technology $\mathcal{A}_0$. Here is a more general formulation that allows for multiple minimal technologies, and also encompasses the simplification in the previous paragraph. Suppose simply that there is some nonempty collection $\mathcal{T}$ of possible technologies, and the principal's value from any contract w is defined as $V_P(w) = \inf_{\mathcal{A} \in \mathcal{T}} V_P(w|\mathcal{A})$. Suppose that $\mathcal{T}$ has the following property: For any $\mathcal{A} \in \mathcal{T}$, and any arbitrary action (F, c), then there exists some $\mathcal{A}' \subseteq \mathcal{A}$ such that $\mathcal{A}' \cup \{(F, c)\} \in \mathcal{T}$. Then, we again have the result that a linear contract is optimal.

The proof is essentially the same as before, using the following generalization of Lemma 2.1: If w is a nonzero contract such that $V_P(w) \geq V_P(0)$, then

$$V_P(w) = \min E_F[y - w(y)] \quad \text{over } F \in \Delta(Y) \text{ such that } E_F[w(y)] \geq \inf_{\mathcal{A} \in \mathcal{T}} V_A(w|\mathcal{A}).$$

If $V_P(w) > 0$, then for any F attaining the minimum, $E_F[w(y)] = \inf_{\mathcal{A}} V_A(w|\mathcal{A})$. (The proof that there exists an optimum is slightly more work than before, but one can derive an analogue to 2.3 and check that it is upper semi-continuous in α, which is enough for existence of the optimum.)

One can also show that linear contracts are uniquely optimal under an appropriate extension of the full-support assumption.

On the other hand, allowing for a wide range of actions to be possible is indispensable. We cannot (for example) restrict attention to technologies that contain only actions "close" to those in the known technology $\mathcal{A}_0$ and expect the same results to hold.

3.3 Lower bounds on cost

Another perhaps unrealistic assumption in the basic model is the lack of any connection between the cost of an action and the output: The agent might potentially have actions that create large amounts of output for free. Indeed, the worst-case action for any contract is one that produces an undesirable distribution F at cost 0. It might be more reasonable to restrict the space of uncertainty, say, by supposing that the principal knows a lower bound on the cost of producing any given level of expected output.

To model this, suppose there is given a convex function $b : \mathbb{R} \to \mathbb{R}^+$, and amend the definition of a technology $\mathcal{A}$ to require that every $(F, c) \in \mathcal{A}$ should satisfy $c \geq b(E_F[y])$. We suppose that the known technology $\mathcal{A}_0$ also satisfies this condition. We again define $V_P(w)$ as the inf of $V_P(w|\mathcal{A})$ over all possible technologies $\mathcal{A} \supseteq \mathcal{A}_0$. It turns out that a linear contract is still optimal.

In fact, a significant generalization holds too. We can allow the known lower bound on cost, b, to depend not only on the expected value of output but also on other moments (for example, it may be that producing higher-variance output can be less costly, whereas deterministically producing the same mean output is known to be expensive). Following Holmström [11], we can also allow there to be other observable variables, besides output, that affect the bound on cost. The general result is that the optimal contract is an affine function of output and whatever other variables affect the bound on cost. However, the argument no longer shows that *all* optimal contracts are affine under the full-support condition.

The argument here is an extension of the ideas used for the basic model, but a rather subtler form of the separation argument is needed. In addition, identifying the worst-case action for a given contract involves a boundary case that previously applied only for the zero contract, but now can occur more widely and so requires more careful treatment. Rather than go through the details here, we defer the statement and proof to Appendix B.

3.4 Risk aversion

The basic model assumed both the principal and agent were financially risk-neutral. This assumption keeps the model as simple as possible, and is particularly convenient for the affine-geometry tools used in the analysis. However, it turns out that the analysis still applies almost identically when the parties have nonlinear utility functions. It would be too much to ask for the optimal contract w to have payments linear in y; instead, we have

linearity in the utility space.

We extend the model as follows. Suppose the principal and agent have increasing, bijective utility functions $u_P, u_A : \mathbb{R} \to \mathbb{R}$. (Note that these conditions imply continuity.) We may normalize $u_P(0) = u_A(0) = 0$. Actions, technologies, and contracts are defined as before, but the payoffs are different. There are two natural specifications for the agent's utility, and we will consider both:

(i) The cost of an action is an additive disutility of effort. In this case, we define $V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}}(E_F[u_A(w(y))] - c)$, and $A^*(w|\mathcal{A})$ is defined as the corresponding argmax.

(ii) The cost is a monetary cost, which the agent directly subtracts from his compensation. Then $V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}}(E_F[u_A(w(y) - c)])$, and $A^*(w|\mathcal{A})$ is the corresponding argmax.

The principal's payoff under $\mathcal{A}$ is defined as $V_P(w|\mathcal{A}) = \max_{(F,c)\in A^*(w|\mathcal{A})} E_F[u_P(y - w(y))]$. As before, the principal's objective is worst-case expected utility, $V_P(w) = \inf_{\mathcal{A}\supseteq\mathcal{A}_0} V_P(w|\mathcal{A})$.

The nontriviality assumption in specification (i) is that there should exist $(F, c) \in \mathcal{A}_0$ with $E_F[u_A(y)] > c$. In specification (ii), we should have $E_F[u_A(y - c)] > 0$. This assumption ensures that the principal can obtain positive expected utility (for example, using a contract $w(y) = \alpha y$ with α close to 1).

We outline the analysis. The zero contract is analyzed as before: If there exists any zero-cost action in $\mathcal{A}_0$ then $V_P(0) = \max_{(F,0)\in\mathcal{A}_0} E_F[u_P(y)]$, and otherwise $V_P(0) = 0$. For other contracts, the worst-case payoff is given by the analogue of Lemma 2.1:

Lemma 3.1. *In the setting with nonlinear utility functions, let w be any nonzero contract such that $V_P(w) \geq V_P(0)$. Then*

$$V_P(w) = \min E_F[u_P(y - w(y))] \quad \text{over } F \in \Delta(Y) \text{ such that } E_F[u_A(w(y))] \geq V_A(w|\mathcal{A}_0).$$

If $V_P(w) > 0$, then for any F attaining the minimum, $E_F[u_A(w(y))] = V_A(w|\mathcal{A}_0)$.

The proof is entirely analogous, with u_P's and u_A's inserted in the relevant places. This holds for either specification of the agent's utility.

To state the main result under nonlinear utility, we say that a contract w is *utility-affine* if there exist constants $\alpha \geq 0$ and β such that $u_A(w(y)) = \alpha u_P(y - w(y)) + \beta$ for all y. The analogue of Theorem 2.3 is then:

Theorem 3.2. *There exists a utility-affine contract that maximizes V_P. If $\mathcal{A}_0$ satisfies the full-support condition, then every contract that maximizes V_P is utility-affine.*

The proof follows that of Theorem 2.3. In the separation step, we now take S to be the convex hull of $\{(u_A(w(y)), u_P(y - w(y))) \mid y \in Y\}$, and take $T = \{(u, v) \mid u > V_A(w|\mathcal{A}_0), v < V_P(w)\}$ just as before. We obtain $\lambda > 0$, $\mu < 0$, and ν such that

$$\lambda u_A(w(y)) + \mu u_P(y - w(y)) \leq \nu$$

for each $y \in Y$, with equality on the support of the worst-case distribution F^*. To construct the new contract w' from w, note that for any value of y, there is a unique value of $w'(y) \geq w(y)$ such that

$$\lambda u_A(w'(y)) + \mu u_P(y - w'(y)) = \nu. \tag{3.3}$$

To see this, treat $w'(y)$ as a variable in (3.3). The left-hand side of (3.3) is continuous and strictly increasing; it is $\leq \nu$ when $w'(y) = w(y)$ and tends to ∞ as $w'(y) \to \infty$, so there is a unique value at which the equality holds. In order to know that w' is a contract, we need to check that it is continuous; this is a straightforward argument (see Appendix C for details). Then, it is clear that w' is utility-affine.

Now essentially the same calculations as before show that $V_P(w'|\mathcal{A}) \geq V_P(w)$, so that $V_P(w') \geq V_P$, and the inequality is strict if $\mathcal{A}_0$ satisfies the full-support condition and $w' \neq w$. All that remains is to check existence of an optimal utility-affine contract. This is a bit more technically involved now than it was under linear utility functions, but presents no new conceptual challenge (again, more details are in Appendix C).

For natural specifications of u_P and u_A, it may typically be impossible to solve the equation $u_A(w(y)) = \alpha u_P(y - w(y)) + \beta$ explicitly for $w(y)$, but it is still possible to deduce qualitative properties of the contract. For example, one can easily show that $w(y)$ is increasing in y (strictly if $\alpha > 0$). If u_P is linear and u_A is concave, then $w(y)$ is convex in y.

We note that Theorem 3.2 states only that the optimal contract is utility-affine, not that it is utility-linear ($\beta = 0$). In the basic model there was a step observing that for any affine contract with $\beta > 0$, we can replace $\beta = 0$ (keeping the same α) and obtain an improvement. This is no longer possible here, because this replacement can change the agent's optimal action under $\mathcal{A}_0$ — and therefore the principal's guarantee from Lemma 3.1 — in unpredictable ways.

3.5 Screening on technology

We note that the maxmin in the principal's problem is generally not equal to the minmax — at least as long as the maxmin-optimal contract is not zero. That is, there exists $\overline{V}_P > \max_w V_P(w)$ such that, if the principal were to know the agent's technology $\mathcal{A}$ at the time of contracting, she could earn an expected payoff of at least $\overline{V}_P$, no matter what the technology $\mathcal{A}$ turned out to be.

This can be seen as follows. Let $(F^*, c^*) \in \mathcal{A}_0$ be the action that maximizes the objective (2.11), $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$, and $w^*(y) = \alpha^* y$ the corresponding maxmin-optimal contract. Assume $\alpha^* > 0$, so that $c^* > 0$. The principal's guarantee is $V_P(w^*) = (\sqrt{E_{F^*}[y]} - \sqrt{c^*})^2$. Consider any technology $\mathcal{A}$, and let (F, c) be the agent's action under w^* and $\mathcal{A}$. Thus

$$\alpha^* E_F[y] - c \geq \alpha^* E_{F^*}[y] - c^*. \tag{3.4}$$

We consider two cases.

- If $c \geq c^*/2$, then the principal's payoff is

$$\begin{aligned}(1-\alpha^*)E_F[y] &\geq \frac{1-\alpha^*}{\alpha^*}\left(\alpha^* E_{F^*}[y] - c^* + c\right) \\ &= E^*_F[y] - 2\sqrt{c^* E_{F^*}[y]} + c^* + \frac{1-\alpha^*}{\alpha^*}c \\ &\geq V_P(w^*) + \frac{1-\alpha^*}{\alpha^*}\frac{c^*}{2}.\end{aligned}$$

- Now suppose $c \leq c^*/2$. We know that if the principal learns $\mathcal{A}$ before contracting, she can earn at least $(\sqrt{E_F[y]} - \sqrt{c})^2$ (since in fact this is her worst-case guarantee with $\mathcal{A}$ in place of $\mathcal{A}_0$ — note the condition $E_F[y] > c$ is met). We compute the minimum of this expression, subject to (3.4) and $c \leq c^*/2$. Define

$$g(x) = \sqrt{\frac{x^2 + (\alpha^* E_{F^*}[y] - c^*)}{\alpha^*}} - x \tag{3.5}$$

  for $x \geq 0$. Then g is convex, and we check that the minimum is given by the first-order condition; this condition is satisfied (uniquely) by $x = \sqrt{c^*}$, with value $g(\sqrt{c^*}) = \sqrt{E_{F^*}[y]} - \sqrt{c^*}$. Now, for any given c, the value of $E_F[y]$ that minimizes $(\sqrt{E_F[y]} - \sqrt{c})^2$, subject to (3.4) and $E_F[y] > c$, is given by taking (3.4) to hold with equality. In this case $E_F[y] = (c + (\alpha^* E_{F^*}[y] - c^*))/\alpha^*$, and so $\sqrt{E_F[y]} - \sqrt{c} = g(\sqrt{c})$.

Thus we see that the principal can make a payoff of at least

$$\begin{aligned}\left(\sqrt{E_F[y]}-\sqrt{c}\right)^2 &\geq \left(g(\sqrt{c})\right)^2 \\ &\geq \left(g(\sqrt{c^*/2})\right)^2 \\ &> \left(g(\sqrt{c^*})\right)^2 \qquad = V_P(w^*).\end{aligned}$$

So in both cases, we have a lower bound for the principal's payoff when she knows $\mathcal{A}$ that is strictly above $V_P(w^*)$.

This is methodologically interesting since it shows that we could not have computed the principal's maxmin payoff by considering a single technology. More importantly, however, it also invites the possibility of screening. That is, the timing of the basic model assumes that the principal can only offer a single contract, without first finding out anything about the agent's technology $\mathcal{A}$ (except $\mathcal{A} \supseteq \mathcal{A}_0$). If the principal could instead ask the agent to announce his $\mathcal{A}$, and choose a contract based on the reported $\mathcal{A}$, could she then guarantee herself a payoff strictly above her maxmin payoff?

It turns out the answer is no, assuming that the screening needs to be incentive-compatible. To formalize this, we imagine that the principal offers a menu of contracts $\mathcal{W} = (w_{\mathcal{A}})$, one for each possible technology $\mathcal{A}$ that the agent could have, such that the agent with any technology $\mathcal{A}$ chooses the corresponding contract (this is without loss of generality by the revelation principle). Thus, we require

$$V_A(w_{\mathcal{A}}|\mathcal{A}) \geq V_A(w_{\mathcal{A}'}|\mathcal{A}) \qquad \text{for all} \qquad \mathcal{A}, \mathcal{A}' \supseteq \mathcal{A}_0. \tag{3.6}$$

We write the principal's worst-case payoff as

$$V_P(\mathcal{W}) = \inf_{\mathcal{A} \supseteq \mathcal{A}_0} V_P(w_{\mathcal{A}}|\mathcal{A}).$$

Theorem 3.3. *The principal cannot do any better, in terms of worst-case guarantee, with a menu of contracts than she can with a single contract. That is, for any menu $\mathcal{W}$,*

$$V_P(\mathcal{W}) \leq \max_w V_P(w).$$

Proof: Consider any menu $\mathcal{W}$. Let $w_0 = w_{\mathcal{A}_0}$, the contract that the agent would choose when the technology is just $\mathcal{A}_0$. We claim that $V_P(w_0) \geq V_P(\mathcal{W})$, which will prove the theorem.

Suppose not. Then, there is some technology $\mathcal{A}_1$ under which, facing contract w_0, the action chooses an action (F_1, c_1) that gives the principal payoff less than $V_P(\mathcal{W})$. We may assume that $\mathcal{A}_1 = \mathcal{A}_0 \cup \{(F_1, c_1)\}$. Note also that $(F_1, c_1) \notin A_0$, since otherwise $\mathcal{A}_1 = \mathcal{A}_0$ and so $V_P(w_0|\mathcal{A}_0) < V_P(\mathcal{W})$ which is a contradiction. It must be that, under w_0, the agent earns strictly higher payoff from (F_1, c_1) than he does from any action in $\mathcal{A}_0$: otherwise he would be willing to take the same action under $\mathcal{A}_1$ as he does under $\mathcal{A}_0$, thereby giving the principal $V_P(w_0|\mathcal{A}_0) \geq V_P(\mathcal{W})$.

Now let $w_1 = w_{\mathcal{A}_1}$, the contract chosen from the menu when the technology is $\mathcal{A}_1$. Under w_1 and $\mathcal{A}_1$, the agent must choose action (F_1, c_1). Proof: If he chooses any action in $\mathcal{A}_0$, then his payoff is at most $V_A(w_0|\mathcal{A}_0)$ (by revealed preference (3.6), since his payoff is the same as $V_A(w_1|\mathcal{A}_0)$). On the other hand, his payoff under w_1 and $\mathcal{A}_1$ must be at least as high as his payoff from (F_1, c_1) under w_0 (by revealed preference again, since w_1 was chosen under $\mathcal{A}_1$), which is higher than $V_A(w_0|\mathcal{A}_0)$ by the previous paragraph.

Hence, (F_1, c_1) is the agent's uniquely chosen action under w_1, and

$$E_{F_1}[w_1(y)] - c_1 \geq E_{F_1}[w_0(y)] - c_1.$$

Then, the principal's payoff when the technology is $\mathcal{A}_1$ is

$$\begin{aligned} E_{F_1}[y - w_1(y)] &= E_{F_1}[y] - c_1 - (E_{F_1}[w_1(y)] - c_1) \\ &\leq E_{F_1}[y] - c_1 - (E_{F_1}[w_0(y)] - c_1) \\ &= E_{F_1}[y - w_0(y)] \\ &< V_P(\mathcal{W}) \end{aligned}$$

where the last line is by definition of (F_1, c_1). Since the principal should get at least $V_P(\mathcal{W})$ under every possible technology, we have a contradiction. □

As a side note, although we showed above that the solution to the principal's maxmin payoff problem is never the solution for any specific technology, one can ask whether the *contract* that solves the maxmin problem is ever optimal for any specific technology. Here the answer is yes. This is shown in Appendix D.

4 Discussion

We have presented here a simple principal-agent model that illustrates the robustness value of linear contracts. In the face of uncertainty about the technology available to the

agent, linearity is the only tool the principal can use to turn her assurance about the agent's expected payoff into a guarantee for herself, and so optimal contracts are linear.

Since one purpose of this paper is to offer a new explanation for the popularity of linear contracts, we should now discuss its relation to the other explanations in the literature. Many previous scholars have noticed that, whereas theoretical models often predict complicated incentive schemes that are sensitive to the details of the model, in practice one often sees simple contracts, and linear contracts are particularly common (see [2, pp. 763-4] and [5, fn. 3] for many references). The model of Holmström and Milgrom [12] quoted above was one early effort to show how the robustness of linear contracts can help explain their popularity. In their model, the principal and agent have CARA utility, and the agent controls the drift of a (possibly multidimensional) Brownian motion in continuous time. Although the principal can condition payments on the entire path of motion, the optimal contract is simply a linear function of the endpoint. Holmström and Milgrom describe this conclusion as a consequence of robustness, in view of the agent's large strategy space. However, it is really the stationary time structure of the model that underlies the conclusion: the CARA utility implies that at each point in time, the optimal incentives going forward are independent of the previous history, and this leads to linearity.

Diamond [7] gives an argument particularly close to the intuition of this paper. Diamond's Section 5 considers a model in which the agent can either choose no effort, producing a low expected output, or high effort, producing a higher expected output. For a given level of effort, the the agent can choose among all distributions over output that have the same mean, and all such distributions are equally costly. A linear contract is then optimal. The argument rests on the same intuition as here — with such freedom to choose the distribution, only a linear bound can tie the principal's expected profit to the agent's expected compensation. However, the assumptions that there are exactly two effort levels, and that all distributions with a given mean are possible, are strong. In any case, there are actually many optimal contracts in Diamond's model. In our model, uncertainty about which distributions are actually possible can make the linear contract *uniquely* optimal.

Laffont and Tirole [14] and McAfee and McMillan [16] consider problems that combine moral hazard and adverse selection: a principal uses a menu of contracts to screen agents on ability. In both of their models, there is an optimal menu in which payment is linear in output within each contract. Again, however, there may also be other optimal menus. Edmans and Gabaix [8] give a general modeling framework that leads to simple, closed-

form contracts; a version of their model with linear utility and additive noise in the contractible outcome leads to linear contracts. However, the assumption of additive noise is restrictive, and that model focuses mainly on implementing a particular action, rather than the more primitive objective of maximizing the principal's payoff. Finally, Hurwicz and Shapiro [13] and Chassang [4, Theorem 1] give maxmin contracting problems with linear solutions; the objective there is the *ratio* of the principal's profit to first-best total surplus. Chassang calculates the worst-case guarantee of linear contracts using the same argument presented at the beginning of Subsection 2.3 of this paper, although the proof of *optimality* of linear contracts is very different. These two papers do not discuss the intuition behind the optimality results, nor do they clarify the motivation for some perhaps unintuitive restrictions on the class of environments considered, which are needed for the results.

Against this backdrop, then, the contribution of our model is a specific combination of features: The model allows many degrees of freedom (the set of *known* actions the agent has can be virtually anything); the concern for robust performance is modeled explicitly through the maxmin payoff objective; and under weak conditions, linear contracts turn out to be *uniquely* optimal.

The mathematical arguments are simple, and this is also a virtue of the model: as discussed in the introduction, one main purpose of the model is to present a methodology that can be adopted to study more complicated contracting problems. The various extensions in Section 3 (and Appendix B) illustrate this. A further illustration is the companion paper [3], which applies the maxmin objective to the principal-expert problem of Zermeño [20, 21] to study worst-case-optimal incentives for information acquisition. Relatedly, the modeling approach here may prove useful to economic theorists developing models of larger phenomena, who need a tractable and flexible model of moral hazard to serve as just one of many moving parts.

A An alternative approach

We give here another approach to the main step of Theorem 2.3: that for any contract w, there is a linear contract w' that guarantees at least as much for the principal. (The argument here was suggested by Lucas Maestri.)

Consider any w with $V_P(w) > 0$, and let (F_0, c_0) be the action that the agent would choose under technology $\mathcal{A}_0$. Put $\alpha = E_{F_0}[w(y)]/E_{F_0}[y]$. (The denominator must be positive, since otherwise the principal is not guaranteed a positive payoff under w.) Put

$w'(y) = \alpha y$. Notice that under this contract, the agent can again take action (F_0, c_0) to earn a payoff of

$$E_{F_0}[\alpha y] - c_0 = E_{F_0}[w(y)] - c_0 = V_A(w|\mathcal{A}_0),$$

and the principal then earns

$$E_{F_0}[(1-\alpha)y] = E_{F_0}[y - w(y)] = V_P(w|\mathcal{A}_0) \geq V_P(w).$$

We will show that the principal does at least as well under w' as under w. Consider an arbitrary technology $\mathcal{A}$, and let (F, c) be the action the agent would take under contract w'; we need to show that the principal's resulting payoff, $V_P(w'|\mathcal{A})$, is at least $V_P(w)$. If $E_F[y] \geq E_{F_0}[y]$, then the principal gets

$$(1-\alpha)E_F[y] \geq (1-\alpha)E_{F_0}[y] = V_P(w|\mathcal{A}_0) \geq V_P(w).$$

Also, we have $E_F[w'(y)] - c \geq V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0)$ by optimality for the agent, and if equality holds throughout, then the agent would also be willing to choose (F_0, c_0), which, again, gives the principal at least $V_P(w)$; thus $V_P(w'|\mathcal{A}) \geq V_P(w)$ in this case too. So we can focus on the case when $E_F[y] < E_{F_0}[y]$ and $E_F[w'(y)] - c > V_A(w|\mathcal{A}_0)$.

Put $\lambda = E_F[y]/E_{F_0}[y]$, and let F' be the mixture $\lambda F_0 + (1-\lambda)\delta_0$. Then, consider contract w when the technology is $\mathcal{A}_0 \cup \{(F', c)\}$. The agent's payoff from (F', c) is

$$\begin{aligned}
E_{F'}[w(y)] - c &= \lambda E_{F_0}[w(y)] + (1-\lambda)w(0) - c \\
&\geq \lambda E_{F_0}[w(y)] - c \\
&= \lambda\alpha E_{F_0}[y] - c \\
&= \alpha E_F[y] - c \\
&= E_F[w'(y)] - c \\
&> V_A(w|\mathcal{A}_0)
\end{aligned}$$

which means that the agent would strictly prefer to take action (F', c) over any other

action. This leaves the principal with a payoff of

$$\begin{aligned} E_{F'}[y - w(y)] &= \lambda E_{F_0}[y - w(y)] - (1-\lambda)w(0) \\ &\leq \lambda E_{F_0}[y - w(y)] \\ &= (1-\alpha)E_F[y] \\ &= E_F[y - w'(y)] \\ &= V_P(w'|\mathcal{A}). \end{aligned}$$

Thus $V_P(w) \leq V_P(w'|\mathcal{A})$ in this case too. So the inequality holds for all $\mathcal{A}$, implying $V_P(w) \leq V_P(w')$.

We comment that, while this proof is quicker and more direct than the separation-based proof in the main text, we have focused on the separation approach for two reasons. One is that that approach generalizes readily, in particular to the multiple-observables extension of Appendix B and to the principal-expert problem in [3]. The approach above depends on taking a convex combination of an arbitrary distribution with δ_0 to attain a specific expected output; it is not clear how to extend it when the space of observable outcomes is not one-dimensional. The second reason is that the second part of Theorem 2.3 — only linear contracts are optimal with full support — is immediate with the separation approach; with the argument here it seems to require extra work.

B General lower bounds on cost

We generalize the basic model to allow for a vector of observable variables $z = (z_1, \ldots, z_k)$, taking values in the compact set $Z \subseteq \mathbb{R}^k$. Thus, an action consists of a distribution over Z and an associated cost. We allow the principal to know a lower bound for the possible cost of producing any distribution, which may depend on the expected values of all the z_i. Thus, we assume given a convex function $b : \mathbb{R}^k \to \mathbb{R}^+$, such that the agent's cost of any distribution F over Z is known to be at least $b(E_F[z])$.

Without loss of generality we can include output y as a component of z, say $y = z_1$, and thus assume $\min\{z_1 \mid z \in Z\} = 0$. (We will let $\underline{z}$ denote an element of Z with $\underline{z}_1 = 0$.) Also, note that this framework allows some components of z to be functions of others. For example, we can represent a situation where only output y is observed, and the principal knows that any distribution F costs at least $h(E_F[y]) - \kappa \cdot Var_F[y]$, where h is some given

convex function. We would capture this by letting

$$Z = \{(y, y^2) \mid y \in Y\}$$

and

$$b(z_1, z_2) = \max\{0, h(z_1) - \kappa(z_2 - z_1^2)\}.$$

Formally, we now define an *action* to be a pair $(F, c) \in \Delta(Z) \times \mathbb{R}^+$ such that $c \geq b(E_F[z])$. A *technology* is a compact set of actions. A technology $\mathcal{A}_0$, the set of known actions, is exogenously given. We make the same nontriviality assumption as before.

A *contract* is a continuous function $w : Z \to \mathbb{R}^+$. The timing of the game is as before. Given contract w and technology $\mathcal{A}$, the agent's utility is $V_A(w|\mathcal{A}) = \max_{(F,c)\in\mathcal{A}} (E_F[w(z)] - c)$ and his choice set is $A^*(w|\mathcal{A}) = \arg\max_{(F,c)\in\mathcal{A}}(E_F[w(z)] - c)$. The principal's expected payoff under $\mathcal{A}$ is $V_P(w|\mathcal{A}) = \max_{(F,c)\in A^*(w|\mathcal{A})} E_F[z_1 - w(z)]$. The principal's objective, $V_P(w)$, is then defined to be the infimum of $V_P(w|\mathcal{A})$ over all technologies $\mathcal{A} \supseteq \mathcal{A}_0$.

The main result, generalizing (the first part of) Theorem 2.3 to this setting, is the following:

Theorem B.1. *There exists a contract that maximizes $V_P(w)$ and is affine — that is, of the form*

$$w(z) = \alpha_1 z_1 + \cdots + \alpha_k z_k + \beta$$

for some real numbers α_i and β.

In the setting in the main body of the paper, where $z = y$, it is easy to check that the optimal affine contract satisfies $0 \leq \alpha_1 < 1$ and $\beta = 0$, so we have the same linearity conclusion as in the basic model.

The proof follows the same outline as in Subsection 2.3. We first characterize the payoff guarantee of any given contract w. The situation is a bit more complex than before, because the assumption of tie-breaking in favor of the principal forces us to deal separately with the boundary case in which the agent's best action under any possible technology is already available in $\mathcal{A}_0$. Previously, this case arose only for the zero contract or other contracts that performed worse, but now it cannot be swept aside so easily.

For $F \in \Delta(Z)$ and a given contract w, define $h(F|w) = E_F[w(z)] - b(E_F[z])$, the highest expected payoff the agent could possibly get from producing distribution F. Since b is convex, h is concave.

Lemma B.2. *Let w be any contract. Then one of the following two cases occurs:*

(i)

$$\begin{aligned} V_P(w) &= \min E_F[z_1 - w(z)] \\ &\quad \text{over } F \in \Delta(Z) \text{ such that } h(F|w) \geq V_A(w|\mathcal{A}_0), \end{aligned} \tag{B.1}$$

and moreover, as long as $V_P(w) > 0$, then for any F attaining the minimum, $h(F|w) = V_A(w|\mathcal{A}_0)$.

(ii)

$$\max_{F \in \Delta(Z)} h(F|w) = V_A(w|\mathcal{A}_0),$$

and

$$V_P(w) = \max E_F[z_1 - w(z)] \quad \text{over } (F, c) \in \mathcal{A}_0 \text{ such that } E_F[w(z)] - c = V_A(w|\mathcal{A}_0). \tag{B.2}$$

Proof: Let F_0 be a distribution attaining the minimum in (B.1). (The constraint set is nonempty since it is satisfied by the action chosen under $\mathcal{A}_0$.) Suppose that F_0 does not also maximize $h(F|w)$ over all $F \in \Delta(Z)$. Then, choose F_1 yielding a higher value of h, and put $F' = (1-\epsilon)F_0 + \epsilon F_1$ for small ϵ. By concavity, $h(F'|w) \geq (1-\epsilon)h(F_0|w) + \epsilon h(F_1|w) > h(F_0|w)$. So if $\mathcal{A} = \mathcal{A}_0 \cup \{(F', b(E_{F'}[z]))\}$, then the agent's unique optimal action in $\mathcal{A}$ is $(F', b(E_{F'}[z]))$. As $\epsilon \to 0$ the principal's resulting payoff tends to $E_{F_0}[z_1 - w(z)]$. Thus the principal cannot be guaranteed more than the value in (B.1). On the other hand the principal is guaranteed at least this much, just as in the proof of Lemma 2.1.

Also, if $h(F_0|w) > V_A(w|\mathcal{A}_0)$ strictly, then let $\mathcal{A} = \mathcal{A}_0 \cup \{(F_0, b(E_{F_0}[z]))\}$. With this technology, the agent's unique optimal action is $(F_0, b(E_{F_0}[z]))$, and again the principal cannot be guaranteed more than the value in (B.1). Thus in either of these situations $V_P(w)$ is as specified by (B.1).

Moreover if $h(F_0|w) > V_A(w|\mathcal{A}_0)$ but $V_P(w) > 0$, then by mixing F_0 with a small point mass on $\underline{z}$ we get a distribution still satisfying the constraint in (B.1) and giving a lower value of the objective, a contradiction. This proves conclusion (i).

We are left with the situation in which F_0 maximizes $h(F|w)$ over all $F \in \Delta(Z)$ and $h(F_0|w) = V_A(w|\mathcal{A}_0)$, so that the first statement of conclusion (ii) holds. In this case, let $(F^*, c^*) \in \mathcal{A}_0$ be the action chosen when the technology is $\mathcal{A}_0$. Then $V_P(w|\mathcal{A}_0)$ equals the maximum in (B.2), attained by (F^*, c^*). This implies that $V_P(w)$ is at most the expression

in (B.2). On the other hand, for any technology $\mathcal{A} \supseteq \mathcal{A}_0$, let (F, c) be the chosen action. We have

$$V_A(w|\mathcal{A}) \leq E_F[w(z)] - b(E_F[z]) = h(F|w) \leq h(F_0|w) = V_A(w|\mathcal{A}_0)$$

and there must be equality throughout. So the agent's expected payoff is always equal to $V_A(w|\mathcal{A}_0)$, and the principal gets at least the maximum in (B.2). Thus, (B.2) is an equality, completing the proof of conclusion (ii). □

Now we prove Theorem B.1 by the same process as before: given a non-affine contract w, use a separation argument to replace it by an affine contract w' that is pointwise above it and gives a weakly greater guarantee to the principal. Whereas the separation argument in the basic model could most conveniently be expressed in payoff space, here we do the separation in outcome space. In addition, we use two different versions of the argument, depending which case of Lemma B.2 applies.

Proof of Theorem B.1: We may assume that the convex hull of Z is a full-dimensional set in $\mathbb{R}^k$. (This can be accomplished by a linear change of coordinates to embed Z in a smaller-dimensional space if necessary, unless $Y = \{0\}$ but the latter situation is uninteresting.)

Consider any non-affine contract w. Nontriviality assures that there exists a contract with positive guarantee, so we may restrict attention to contracts with $V_P(w) > 0$. One of the two cases of Lemma B.2 holds, and we deal with the two separately.

Case (i). We define

$$t(z) = \max\{b(z), z_1 - V_P(w) - V_A(w|\mathcal{A}_0)\}$$

and observe that t is a convex function. Now, we define two sets in $\mathbb{R}^{k+1} = \mathbb{R}^k \times \mathbb{R}$. Let S be the convex hull of all pairs $(z, w(z) - V_A(w|\mathcal{A}_0))$. Let T be the set of all pairs (z, c) such that z lies in the convex hull of Z, and $c > t(z)$.

Both of these sets are convex. We claim they are disjoint. If not, there exists some $F \in \Delta(z)$ such that

$$E_F[w(z)] - V_A(w|\mathcal{A}_0) > t(E_F[z]).$$

In particular,

$$E_F[w(z)] - V_A(w|\mathcal{A}_0) > E_F[z_1] - V_P(w) - V_A(w|\mathcal{A}_0)$$

implying

$$E_F[z_1 - w(z)] < V_P(w),$$

and also

$$E_F[w(z)] - V_A(w|\mathcal{A}_0) > b(E_F[z])$$

implying

$$h(F|w) > V_A(w|\mathcal{A}_0).$$

This is a direct contradiction to our statement (i).

So by the separating hyperplane theorem, there are constants $\lambda_1, \ldots, \lambda_k, \mu, \nu$ such that

$$\sum_i \lambda_i z_i + \mu c \leq \nu \qquad \text{for all} \qquad (z, c) \in S, \tag{B.3}$$

$$\sum_i \lambda_i z_i + \mu c \geq \nu \qquad \text{for all} \qquad (z, c) \in T, \tag{B.4}$$

and some λ_i or μ is nonzero. Inequality (B.4) implies $\mu \geq 0$. In fact, $\mu > 0$. Proof: Suppose $\mu = 0$. Since the projection of either S or T onto the first k coordinates contains Z, (B.3) gives $\sum_i \lambda_i z_i \leq \nu$ for all $z \in Z$, while (B.4) gives $\sum_i \lambda_i z_i \geq \nu$ for all $z \in Z$. Hence, $\sum_i \lambda_i z_i = \nu$ for all $z \in Z$. Since not all λ_i are zero, this contradicts the full-dimensionality of Z.

We can rewrite (B.3) as

$$w(z) \leq \frac{\nu - \sum_i \lambda_i z_i}{\mu} + V_A(w|\mathcal{A}_0) \qquad \text{for all } z \in Z.$$

This motivates us to define

$$w'(z) = \frac{\nu - \sum_i \lambda_i z_i}{\mu} + V_A(w|\mathcal{A}_0),$$

an affine contract satisfying $w' \geq w$ pointwise.

Also, let (F_0, c_0) be the action chosen by the agent under w and technology $\mathcal{A}_0$. Which of the two branches of t occurs at $E_{F_0}[z]$? Observe that

$$E_{F_0}[z_1] - V_P(w) - V_A(w|\mathcal{A}_0) \geq E_{F_0}[z_1] - V_P(w|\mathcal{A}_0) - V_A(w|\mathcal{A}_0) = c_0 \geq b(E_{F_0}[z]),$$

hence $t(E_{F_0}[z]) = E_{F_0}[z_1] - V_P(w) - V_A(w|\mathcal{A}_0)$. Thus we conclude from (B.4) that

$$E_{F_0}[z_1] - V_P(w) - V_A(w|\mathcal{A}_0) \geq \frac{\nu - \sum_i \lambda_i E_{F_0}[z_i]}{\mu}. \tag{B.5}$$

Now we are ready to check that $V_P(w') \geq V_P(w)$. Certainly, we have $V_A(w'|\mathcal{A}_0) \geq V_A(w|\mathcal{A}_0)$, since whichever action the agent takes under w and $\mathcal{A}_0$ gives him a weakly higher payoff under w'. Since the agent can only do better under any larger technology $\mathcal{A}$ than $\mathcal{A}_0$, then actually $V_A(w'|\mathcal{A}) \geq V_A(w|\mathcal{A}_0)$.

Suppose that for some technology $\mathcal{A} \supseteq \mathcal{A}_0$, the agent takes action (F, c). Then (B.4) implies

$$\begin{aligned} t(E_F[z]) &\geq \frac{\nu - \sum_i \lambda_i E_F[z_i]}{\mu} \\ &= E_F[w'(z)] - V_A(w|\mathcal{A}_0) \\ &= V_A(w'|\mathcal{A}) + c - V_A(w|\mathcal{A}_0) \\ &\geq c \\ &\geq b(E_F[z]). \end{aligned}$$

If the inequality is strict, then $t(E_F[z]) = E_F[z_1] - V_P(w) - V_A(w|\mathcal{A}_0)$ and so we have

$$V_P(w'|\mathcal{A}) = E_F[z_1 - w'(z)] = t(E_F[z]) + V_P(w) + V_A(w|\mathcal{A}_0) - E_F[w'(z)] \geq V_P(w).$$

Otherwise, $t(E_F[z]) = b(E_F[z])$ and so all the inequalities in the stacked chain above are equalities. In particular, the second inequality is an equality, implying $V_A(w'|\mathcal{A}) = V_A(w'|\mathcal{A}_0) = V_A(w|\mathcal{A}_0)$. Since the agent does at least as well as $V_A(w|\mathcal{A}_0)$ by taking action (F_0, c_0), this action is in his choice set under w' and $\mathcal{A}$, and so the principal gets at least the corresponding payoff:

$$V_P(w'|\mathcal{A}) \geq E_{F_0}[z_1 - w'(z)] = E_{F_0}[z_1] - \frac{\nu - \sum_i \lambda_i E_{F_0}[z_i]}{\mu} - V_A(w|\mathcal{A}_0) \geq V_P(w)$$

from (B.5). Thus in either case, $V_P(w'|\mathcal{A}) \geq V_P(w)$. This holds for all $\mathcal{A}$, so $V_P(w') \geq V_P(w)$.

Case (ii). In this case, define S to be the convex hull of all pairs $(z, w(z) - V_A(w|\mathcal{A}_0))$, and T to be the set of all (z, c) with z in the convex hull of Z and $c > b(z)$. These are convex, and disjoint: otherwise, there exists F such that

$$E_F[w(z)] - V_A(w|\mathcal{A}_0) > b(E_F[z])$$

which simplifies to

$$h(F|w) > V_A(w|\mathcal{A}_0),$$

in contradiction to the statement of (ii). Using the same arguments as in case (i), we find $\lambda_1, \ldots, \lambda_k, \mu, \nu$ such that

$$\sum_i \lambda_i z_i + \mu c \leq \nu \qquad \text{for all} \qquad (z, c) \in S, \tag{B.6}$$

$$\sum_i \lambda_i z_i + \mu c \geq \nu \qquad \text{for all} \qquad (z, c) \in T, \tag{B.7}$$

and we show that $\mu > 0$. Again, (B.6) implies

$$w(z) \leq \frac{\nu - \sum_i \lambda_i z_i}{\mu} + V_A(w|\mathcal{A}_0) \qquad \text{for all } z \in Z.$$

Define $w'(z)$ as the right side of this inequality, so that we have an affine contract satisfying $w' \geq w$ pointwise.

Consider the agent's behavior under contract w'. For any action (F, c) chosen by the agent under any possible technology, we have

$$E_F[w'(z)] - c \leq E_F[w'(z)] - b(E_F[z]) = w'(E_F[z]) - b(E_F[z]) \leq V_A(w|\mathcal{A}_0)$$

where the second inequality follows from (B.7). That is, the agent can never earn a higher expected payoff than $V_A(w|\mathcal{A}_0)$. On the other hand, the agent can always earn at least this much, since

$$V_A(w'|\mathcal{A}) \geq E_{F_0}[w'(z)] - c_0 \geq E_{F_0}[w(z)] - c_0 = V_A(w|\mathcal{A}_0) \tag{B.8}$$

where (F_0, c_0) is his action under w and technology $\mathcal{A}_0$. So we have equality throughout in (B.8). Then the agent's choice set under w' and any technology $\mathcal{A} \supseteq \mathcal{A}_0$ always includes (F_0, c_0), so the principal gets at least $E_{F_0}[z_1 - w'(z)]$. From (B.8), this is equal to $E_{F_0}[z_1 - w(z)]$. But the latter is simply equal to $V_P(w)$ by (ii). So under any technology, the principal gets at least $V_P(w)$ under contract w'.

Existence of an optimum. We have shown that any contract w with $V_P(w) > 0$ can be (weakly) improved to an affine contract. So it now suffices to show existence of an optimum within the class of affine contracts, and this contract will then be optimal among all contracts.

Put $\overline{b} = \max_{z \in Z} b(z)$ and $\overline{y} = \max(Y)$. Note that for any contract w satisfying $\max_{z \in Z} w(z) - \overline{b} \geq \overline{y}$, the agent can potentially attain a payoff greater than $\overline{y}$, which

means that the principal cannot be guaranteed a positive payoff. Hence we can restrict attention to contracts with $w(z) \in [0, \overline{y} + \overline{b}]$ for all z. By full-dimensionality, this implies a compact range of possible values for α and β. We will show below that $V_P(w)$ is upper semi-continuous with respect to w, under the sup-norm topology on the space of contracts. (It may not be fully continuous.) Since the affine contract w in turn varies continuously in α, β under this topology, it will then follow that $V_P(w)$ is upper semi-continuous in α, β, so that the maximum is attained.

Let $w_1, w_2, \ldots$ be any contracts that converge to some contract w_∞ in the sup norm. We wish to show that $V_P(w_\infty) \geq \limsup_k V_P(w_k)$. We can replace the sequence (w_k) with a subsequence along which $V_P(w_k)$ converges to its lim sup on the original sequence; thus, we assume henceforth that $V_P(w_k)$ converges. Now consider any technology $\mathcal{A}$, and let (F_k, c_k) be the agent's chosen action under $\mathcal{A}$ and contract w_k. We may again pass to a subsequence and assume that (F_k, c_k) has some limit $(F_\infty, c_\infty) \in \mathcal{A}$. Then straightforward continuity arguments show that (F_∞, c_∞) is an optimal action (perhaps not the only one) for the agent under w_∞, and its payoff to the principal is the limit of the corresponding payoffs of (F_k, c_k) under w_k. Hence,

$$V_P(w_\infty|\mathcal{A}) \geq E_{F_\infty}[z_1 - w_\infty(z)] = \lim_k E_{F_k}[z_1 - w_k(z)] = \lim_k V_P(w_k|\mathcal{A}) \geq \lim_k V_P(w_k),$$

and so $V_P(w_\infty) \geq \lim_k V_P(w_k)$ as needed. □

Note that the full-support condition does not in general ensure that the affine contract w' is a strict improvement over w for the principal, in either case (i) or (ii). This is because, with multiple observables, an affine contract no longer ties the principal's payoff directly to the agent's; we need to use other arguments to show that $V_P(w') \geq V_P(w)$. So even though the agent does strictly better under w' than w (under full support), we can no longer leverage this fact to show that the principal also does strictly better and conclude that affine contracts are uniquely optimal.

C Detailed arguments with risk-aversion

We avoid going through every step of Theorem 3.2 in full detail, but it is necessary to describe some of the continuity arguments that are more technically involved than their counterparts in the basic model.

To know that the function $w'(y)$ defined by (3.3) is a contract, we need to check that it is continuous. In fact we show more. For *every* real number y, and all $\lambda > 0$, $\mu \leq 0$,

and $\nu \in \mathbb{R}$, define $w'(y;\lambda,\mu,\nu)$ uniquely by

$$\lambda u_A(w'(y;\lambda,\mu,\nu)) + \mu u_P(y - w'(y;\lambda,\mu,\nu)) = \nu. \tag{C.1}$$

We check that $w'(y)$ is *jointly* continuous in y and the parameters λ, μ, ν. Indeed: within any compact region of (y,λ,μ,ν)-space, (C.1) implies $\underline{w} \leq w'(y) \leq \overline{w}$ for some bounds $\underline{w}, \overline{w}$. Now if we take a sequence $(y_k,\lambda_k,\mu_k,\nu_k) \to (y,\lambda,\mu,\nu)$ in this space, such that $w'(y_k;\lambda_k,\mu_k,\nu_k) \not\to w'(y;\lambda,\mu,\nu)$, then compactness implies that there is some subsequence along which $w'(y_k;\lambda_k,\mu_k,\nu_k)$ converges to some value $\widehat{w}' \neq w'(y;\lambda,\mu,\nu)$. Then, by continuity, (C.1) holds at $(y;\lambda,\mu,\nu)$ for both $w'(y;\lambda,\mu,\nu)$ and $\widehat{w}'$, which is impossible.

The other technical step involves checking existence of an optimal contract. Since every contract is weakly outperformed by a utility-affine contract, it suffices to show that among the utility-affine contracts there is one that is optimal. Writing $u_A(w(y)) = \alpha u_P(y - w(y)) + \beta$, we can restrict to a compact set of pairs (α,β). (For example, we can restrict to all contracts satisfying $0 \leq w(y) \leq C$ for all y and sufficiently large constant C, since otherwise there is some technology under which the agent can obtain a payoff larger than C and thereby force the principal's payoff below zero. It is straightforward to check that this restriction, together with $u_P(0 - w(0)) \leq 0$ while $\max_y u_P(y - w(y))$ is bounded above 0, implies a compact set of possible pairs (α,β).) The principal's guarantee, $V_P(w)$, is in turn upper semi-continuous in w under the sup-norm topology (see the end of the proof of Theorem B.1 in Appendix B), and the joint continuity result of the previous paragraph then implies that utility-affine contract w is continuous (under this same topology) in the parameters α, β. Therefore, the optimum exists. B.

D Optimizing for a specific technology

Given the known technology $\mathcal{A}_0$, let $(F^*, c^*) \in \mathcal{A}_0$ maximize (2.11), and $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$, so that $w^*(y) = \alpha^* y$ is the maxmin-optimal contract. We show here that there is a specific technology $\mathcal{A}$ such that this contract w^* is also optimal when the technology is known to be $\mathcal{A}$.

Note that under w^*, when the technology is $\mathcal{A}_0$, the action (F^*, c^*) does in fact maximize the agent's expected payoff $E_F[w^*(y)] - c = \alpha^* E_F[y] - c$. Indeed, this follows from our observations about the function $g(x)$ defined in (3.5); recall that the minimum value

of g was $\sqrt{E_{F^*}[y]} - \sqrt{c^*}$. If some other $(F, c) \in \mathcal{A}_0$ satisfies

$$\alpha^* E_F[y] - c > \alpha^* E_{F^*}[y] - c^*,$$

then

$$\sqrt{E_F[y]} - \sqrt{c} > g(\sqrt{c}) \geq \sqrt{E_{F^*}[y]} - \sqrt{c^*} \tag{D.1}$$

which contradicts the definition of (F^*, c^*).

Now choose some sufficiently high cost limit $\overline{c}$, and let $\mathcal{A}$ be the set of all actions $(F, c) \in \Delta(Y) \times [0, \overline{c}]$ satisfying $\alpha^* E_F[y] - c \leq \alpha^* E_{F^*}[y] - c^*$. It is clear that this is a technology (i.e. it is compact), and by the preceding paragraph, it contains $\mathcal{A}_0$. Under this technology, if the principal offers contract w^*, then the agent is indifferent among all actions on the frontier $\alpha^* E_F[y] - c = \alpha^* E_{F^*}[y] - c^*$. Hence, the agent uses the best such action for the principal, which has $F = \delta_{\overline{y}}$ (a point mass on $\overline{y} = \max(Y)$), and the principal's resulting payoff is $(1 - \alpha^*)\overline{y}$.

We would like to show that no other contract w can deliver a higher payoff under this technology. Suppose the principal offers w, and the agent chooses action (F, c). We have two cases:

- If $E_F[y] \leq (1 - \alpha^*)E_{F^*}[y]$, then clearly the principal's payoff is at most $E_F[y] \leq (1 - \alpha^*)E_{F^*}[y] \leq (1 - \alpha^*)\overline{y}$.

- Otherwise, let F' be a mixture of F and δ_0, with weight $(1 - \alpha^*)E_{F^*}[y]/E_F[y]$ on F and the remaining weight on δ_0. We claim that $(F', 0) \in \mathcal{A}$. Indeed:

  $$\alpha^* E_{F'}[y] - 0 = \alpha^* \frac{(1 - \alpha^*)E_{F^*}[y]}{E_F[y]} E_F[y] = \alpha^* E_{F^*}[y] - \alpha^{*2} E_{F^*}[y] = \alpha^* E_{F^*}[y] - c^*.$$

  Hence, the agent must be compensated enough under (F, c) to prefer this action over $(F', 0)$:

  $$E_F[w(y)] - c \geq E_{F'}[w(y)] \geq \frac{(1 - \alpha^*)E_{F^*}[y]}{E_F[y]} E_F[w(y)],$$

  from which

  $$E_F[w(y)] \geq \frac{E_F[y]}{E_F[y] - (1 - \alpha^*)E_{F^*}[y]} c.$$

Combining with $c \geq \alpha^*(E_F[y] - E_{F^*}[y]) + c^*$ gives

$$E_F[w(y)] \geq E_F[y] \cdot \frac{\alpha^*(E_F[y] - E_{F^*}[y]) + c^*}{E_F[y] - E_{F^*}[y] + \alpha^* E_{F^*}[y]} = \alpha^* E_F[y]$$

(the fraction simplifies to α^* when we recall $\alpha^* = \sqrt{c^*/E_{F^*}[y]}$). Therefore, the principal's payoff is

$$E_F[y - w(y)] \leq (1 - \alpha^*)E_F[y] \leq (1 - \alpha^*)\overline{y}.$$

This shows that no other contract can do better than w^* for the principal, as claimed.

References

[1] Dirk Bergemann and Stephen Morris (2005), "Robust Mechanism Design," *Econometrica* 73 (6), 1771-1813.

[2] Sugato Bhattacharyya and Francine Lafontaine (1995), "Double-Sided Moral Hazard and the Nature of Share Contracts," *RAND Journal of Economics* 26 (4), 761-781.

[3] Gabriel Carroll (2012), "Robust Incentives for Information Acquisition," in preparation.

[4] Sylvain Chassang (2011), "Calibrated Incentive Contracts," working paper, Princeton University.

[5] Leon Yang Chu and David E. M. Sappington (2007), "Simple Cost-Sharing Contracts," *American Economic Review* 97 (1), 419-428.

[6] Kim-Sau Chung and J. C. Ely (2007), "Foundations of Dominant-Strategy Mechanisms," *Review of Economic Studies* 74 (2), 447-476.

[7] Peter Diamond (1998), "Managerial Incentives: On the Near Linearity of Optimal Compensation," *Journal of Political Economy* 106 (5), 931-957.

[8] Alex Edmans and Xavier Gabaix (2011), "Tractability in Incentive Contracting," *Review of Financial Studies* 24 (9), 2865-2894.

[9] Alexander Frankel (2011), "Aligned Delegation," working paper, University of Chicago Booth School of Business.

[10] Daniel Garrett (2012), "Robustness of Simple Menus of Contracts in Cost-Based Procurement," working paper, Toulouse School of Economics.

[11] Bengt Holmström (1979), "Moral Hazard and Observability," *Bell Journal of Economics* 10 (1), 74-91.

[12] Bengt Holmström and Paul Milgrom (1987), "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55 (2), 303-328.

[13] Leonid Hurwicz and Leonard Shapiro (1978), "Incentive Structures Maximizing Residual Gain under Incomplete Information," *Bell Journal of Economics* 9 (1), 180-191.

[14] Jean-Jacques Laffont and Jean Tirole (1986), "Using Cost Observation to Regulate Firms," *Journal of Political Economy* 94 (3), 614-641.

[15] Eric Maskin and Tomas Sjöström (2002), "Implementation Theory," in *Handbook of Social Choice and Welfare*, vol. 1, ed. Kenneth J. Arrow, A. K. Sen, and Kotaro Suzumura (Amsterdam: North-Holland), 237-288.

[16] R. Preston McAfee and John McMillan (1987), "Competition for Agency Contracts," *RAND Journal of Economics* 18 (2), 296-307.

[17] Silvio Micali and Paul Valiant (2008), "Resilient Mechanisms for Truly Combinatorial Auctions," working paper, University of California at Berkeley.

[18] Ilya Segal (2003), "Optimal Pricing Mechanisms with Unknown Demand," *American Economic Review* 93 (3), 509-529.

[19] Takuro Yamashita (2012), "A Necessary Condition for Implementation in Undominated Strategies, with Applications to Robustly Optimal Trading Mechanisms," working paper, Toulouse School of Economics.

[20] Luis Zermeño (2011), "A Principal-Expert Model and the Value of Menus," working paper, Massachusetts Institute of Technology.

[21] Luis Zermeño (2012), "The Role of Authority in a General Principal-Expert Model," working paper, Massachusetts Institute of Technology.