Robustness and Linear Contracts

Gabriel Carroll, Stanford University
gdc@stanford.edu
September 3, 2013

Abstract

We consider a simple moral hazard problem in which the principal is uncertain about what the agent can and cannot do. The principal knows some actions available to the agent, but other, unknown actions may also exist. The principal wants to write a contract that is robust to this uncertainty, and so evaluates contracts by their worst-case performance, with respect to unknown actions the agent might potentially take. The model assumes risk-neutrality and limited liability, and makes no other functional form assumptions. Under very general circumstances, the optimal contract is linear. The model thus offers a new explanation for the widespread use of linear contracts in practice. It also introduces a flexible and tractable modeling approach for moral hazard under non-quantifiable uncertainty; this is underscored by several extensions that show how the basic argument generalizes.

Keywords: limited liability, linear contracts, principal-agent problem, robustness, worst-case

Thanks to (in random order) Larry Blume, Daron Acemoglu, Luis Zermeño, Paul Milgrom, Drew Fudenberg, Iván Werning, Alex Wolitzky, Mike Riordan, Lucas Maestri, David Easley, Ben Golub, Alex Frankel, Sylvain Chassang, Abhijit Banerjee, Bengt Holmström, and Tomasz Sadzik for helpful discussions and advice, as well as other seminar participants at Zurich, UIUC, NYU, UCLA, Toulouse, Caltech, Wisconsin, and Harvard EconCS for their comments.
1 Introduction

This paper considers a simple principal-agent problem with uncertain technology. As in the standard 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. This paper departs from most of the principal-agent literature, however, in how it models the principal’s knowledge of the agent’s possible actions (and their costs). Rather than assuming that this set of actions is known, as is usually done, 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 seeks robustness, and so evaluates possible contracts using a worst-case criterion: any contract is judged 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.

Linear contracts — which pay the agent a fixed share of output — are one way to obtain a worst-case payoff guarantee. For example, suppose the principal considers a contract that pays the agent one-third of output, keeping two-thirds for herself, and suppose she knows some action the agent can take that would give him an expected payoff of 400 under this contract. Then, any unknown action the agent might rationally choose would also give him at least 400. Since the principal’s ex-post payoff is always at least twice the agent’s, she thus is guaranteed at least 800. The main point of this paper is that in fact the best such guarantee, out of all possible contracts, comes from a linear contract. This result holds without any assumptions on the structure of the set of known actions. It also persists — with suitable modifications — through a number of extensions and variations of the basic model.

Briefly, the intuition behind the optimality of linear contracts 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 (from the actions that are known to be available). The only way to turn this into a lower bound on her own expected payoff is to impose a linear relationship between the two, as in the example above. A nonlinear contract may also offer a payoff guarantee, but whenever this happens, the guarantee is still driven by a linear relationship (in general, an inequality).

The importance of our 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. As Holmström and Milgrom write in their classic paper on linear contracts in dynamic environments [15, 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.

This paper aims to answer their implicit call with a forthrightly non-Bayesian model of robustness. 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. This makes it all the more surprising that it did not appear much earlier in the agency theory literature. There have been results on optimality of linear contracts using other maxmin-type criteria, due to Hurwicz and Shapiro [16] and recently Chassang [4, Corollary 1]. Diamond [8] also gave a Bayesian model in which related intuitions apply. However, the present paper offers a relatively general class of environments, together with a mathematical argument for robustness based on the alignment between the principal’s and agent’s goals, that differentiate it from previous literature. (The connections with these previous works will be discussed in more detail in the concluding section, since the discussion will be easier after the model has been presented.) Accordingly, this paper aims to fill a longstanding gap in the arsenal of tools for studying agency problems. From this point of view, its simplicity is a strength, since there is room to expand the method to apply to more complex problems.

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 either remedy unrealistic features of the basic model or otherwise enrich it. This includes assuming some knowledge about the costs of various actions, allowing a participation constraint, and incorporating risk-aversion, as well as allowing the principal to screen agents on their technology. These extensions also illustrate how the method extends beyond the basic model.
This paper joins a recently growing literature exploring mechanism design with worst-case objectives. This includes the work of Hurwicz and Shapiro [16] mentioned above, Frankel [11], and Garrett [12], also on contracting with unknown agent preferences; the work initiated by Bergemann and Morris [1] and Chung and Ely [7] on mechanism design with unknown higher-order beliefs; and work such as Yamashita’s [25] on maxmin expected welfare under weak assumptions on agent behavior (in this case, assuming only that agents play undominated strategies). The implementation literature (surveyed in [19]) 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 the work of Chassang [4], Segal [22], Chawla, Hartline, Malec, and Sivan [5], and Micali and Valiant [21]. There is also a less closely related strand of literature, such as Madarász and Prat [18], that looks at local robustness when the model of the environment is slightly misspecified.

This paper also adds to the literature on explanations for linear contracts — including the maxmin-optimality papers mentioned above as well as several others. Again, discussion of the relationship to that literature is deferred to the concluding section. The conclusion also gives some discussion of interpretation and how to connect the stark assumptions of the model to real-world contract design.

2 The basic model

We start with the basic version of the model. The model here is kept simple, at some costs of realism, which will be addressed later.

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 $x \in X$, $\delta_x$ 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,
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 \( \min(Y) = 0 \).

To model the agent’s actions, we abstract away from their physical description and record only the features that affect behavior and payoffs: the cost of each action to the agent, and the resulting probability distribution over output. Thus, an action is a pair \((F, c) \in \Delta(Y) \times \mathbb{R}^+\). The interpretation is that the agent pays cost \( c \), and output is drawn \( y \sim F \). \( c \) 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 \( \Delta(Y) \times \mathbb{R}^+ \), describing a possible set of actions available to the agent. The agent has a technology \( \mathcal{A} \), which he knows but the principal does not. Instead, the principal knows only some set \( \mathcal{A}_0 \) of actions available to the agent, and she believes \( \mathcal{A} \) 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 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 \( Y \). Our main result becomes stronger when this condition holds.

Next we define the space of contracts. Any contract must 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 \( w : Y \to \mathbb{R}^+ \).

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

\(^1\)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.
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

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

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

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

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

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

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 subsection, we will show that the maximum exists, and identify the contract that attains it.\(^2\)

### 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 \( \alpha \) — can provide this guarantee. The argument was sketched in the introduction, and now we write it out formally. (This same calculation appears also in Chassang [4, Theorem 1].) Suppose the principal offers such a contract, with \( 0 < \alpha \leq 1 \). Note that whatever

---

\(^2\)A referee has pointed out a peculiar feature of the setup here: in the standard benchmark, i.e. where the technology is known to be \( \mathcal{A}_0 \) and so the principal maximizes \( V_P(w|\mathcal{A}_0) \), our assumptions are not sufficient to ensure that an optimal contract exists. This is not really a problem; if one is troubled by nonexistence of the optimum in the benchmark, the usual response would be to impose additional assumptions to ensure existence (such as making \( Y \) finite), and of course our results then continue to be true.
technology $\mathcal{A} \supseteq \mathcal{A}_0$ the agent has, and whatever optimal action $(F, c)$ he chooses,
\[
\alpha \cdot E_F[y] \geq E_F[w(y)] - c = V_A(w|\mathcal{A}) \geq V_A(w|\mathcal{A}_0).
\]

Here the second inequality holds because $V_A(w|\mathcal{A})$ represents the best payoff from $\mathcal{A}$, which is 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,
\[
V_P(w) \geq \frac{1 - \alpha}{\alpha} V_A(w|\mathcal{A}_0) = \max_{(F,c) \in \mathcal{A}_0} \left( (1 - \alpha) E_F[y] - \frac{1 - \alpha}{\alpha} c \right).
\]

The nontriviality assumption implies that if $\alpha$ 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 here is 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 $V_P(w)$. 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, every distribution $F$ over outcomes satisfying $E_F[w(y)] \geq 123$ must also satisfy $E_F[y - w(y)] \geq 456$, otherwise the principal would not be guaranteed 456.

Applying a separation theorem, we conclude that there exist constants $\kappa, \lambda$ such that $y - w(y) \geq \kappa w(y) + \lambda$ for all $y$, and $\kappa \cdot 123 + \lambda \geq 456$.

Now we construct $w'$ such that $y - w'(y) = \kappa w'(y) + \lambda$, 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 (aside from an additive constant term, which is easily dealt with). 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, directly constructive way to show that any contract is (weakly)
outperformed by a linear contract. That alternative proof is a little faster, but generalizes less readily. See Appendix A for details and discussion.

The first step of our argument is to exactly identify the guarantee $V_P(w)$ from any given contract $w$. The characterization (Lemma 2.1 below) is intuitive: minimize the principal’s expected payoff over all output distributions, subject to 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. We include the proof here for completeness, but it can be skipped on a first reading.

One technicality 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, $A^*(0|\mathcal{A}) = \{(F, c) \in \mathcal{A} | c = 0\}$; 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,

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

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

**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, his chosen action $(F, c)$ satisfies

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

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 $\delta_{y^*}$ with weight $\epsilon$, 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 \( A \) 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 \( E_F[w(y)] > V_A(w|A_0) \), then we can again proceed as above with \( A = A_0 \cup \{(F, 0)\} \). This leaves only the case of equality — \( V_A(w|A_0) = \max_y w(y) \) — which is only satisfied when \( A_0 \) contains some action of the form \((F, 0)\) with \( F \) supported at output levels for which \( w \) attains its maximum. Then, under technology \( A_0 \) the agent must choose such an action. Then the principal’s expected payoff \( V_P(w|A_0) \) is the maximum of \( E_F[y] - \max_y w(y) \) over all such actions \((F, 0)\). But this is less than \( E_F[y] \leq \max_{(F, 0) \in A_0} E_F[y] = V_P(0) \). 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 \( E_F[w(y)] > V_A(w|A_0) \) strictly, then replace \( F \) by a mixture of \( F \) with weight \( 1 - \epsilon \) and \( \delta_0 \) with weight \( \epsilon \), for small \( \epsilon \), to see that minimality is contradicted. Hence we have equality, \( E_F[w(y)] = V_A(w|A_0) \), as claimed. \( \square \)

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 \( w(y) = \alpha y \) satisfies \( V_P(w) \geq V_P(0) \) and \( V_P(w) > 0 \), then

\[
V_P(w) = \max_{(F, c) \in A_0} \left( (1 - \alpha)E_F[y] - \frac{1 - \alpha}{\alpha}c \right).
\]

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 \( A_0 \) 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 \( V_P(w) > 0 \). Assume that \( w \) is not linear. 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.

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 \( u > V_A(w|A_0) \) and \( v < V_P(w) \). The conclusion (2.2) of Lemma 2.1 implies that \( S \) and \( T \) are disjoint. (In words, \( S \) is the set of expected (agent,principal)-payoff pairs — ignoring the costs of action — that can occur, while \( T \) consists of pairs that will not occur because they would contradict the principal’s guarantee.) So by the separating hyperplane theorem, there exist constants \( \lambda, \mu, \nu \) such that

\[
\lambda u - \mu v \leq \nu \quad \text{for all} \quad (u, v) \in S, \tag{2.4}
\]

\[
\lambda u - \mu v \geq \nu \quad \text{for all} \quad (u, v) \in T, \tag{2.5}
\]

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 \( \lambda, \mu \geq 0 \). Let us show that both these inequalities hold strictly. If \( \mu = 0 \), then \( \lambda > 0 \), and (2.4) and (2.5) imply \( \max_{y \in Y} w(y) \leq \nu/\lambda \leq V_A(w|A_0) \). 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 the new contract by

\[
w'(y) = \frac{\nu + \mu y}{\lambda + \mu}.
\]

Thus \( w' \geq w \) pointwise. Notice that this immediately implies \( w'(y) \geq 0 \) for all \( y \), so that \( w' \) 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
\[
E_F[w'(y)] \geq E_F[w'(y)] - c = V_A(w'|\mathcal{A}) \geq V_A(w'|\mathcal{A}_0). \tag{2.7}
\]

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

The last inequality holds because \( w' \geq w \) pointwise, so that any action in \( \mathcal{A}_0 \) gives the agent at least as high utility under \( w' \) as \( w \). Moreover, if the full-support condition holds, and \( w' \) does not coincide with \( w \), then that last inequality is strict. (Let \( (F, c) \) be the action taken under \( w \) and technology \( \mathcal{A}_0 \). Since \( V_P(w) > 0 \), we cannot have \( (F, c) = (\delta_0, 0) \), therefore \( F \) has full support. Hence \( E_F[w'(y)] - c > E_F[w(y)] - c \).)

Since (2.10) holds for all \( \mathcal{A} \), we conclude that \( V_P(w') \geq V_P(w) \), with strict inequality if the full-support condition holds and \( w' \neq w \).

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 \( w' \neq w \). Moreover, \( w' \geq w \) pointwise, and in particular \( \beta = w'(0) \geq 0 \). Now replace \( w'(y) \) by \( \alpha y = w'(y) - \beta \); this further increases \( V_P(w') \) by \( \beta \), 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 \( w \). Moreover, since we assumed \( w \) was not already linear, at least one of our steps actually changed the contract, and under the full-support condition this implies a strict improvement.

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), which was a lower bound for the guarantee of the linear contract with share \( \alpha \), with equality whenever the true guarantee is both \( > 0 \) and \( \geq V_P(0) \) (which must be true for some linear contract). Since (2.3) is continuous in the share \( \alpha \in [0, 1] \), it achieves a maximum, and therefore this maximum is also the optimal guarantee over all 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.

Although Theorem 2.3 was our main goal, we can easily complete the analysis of the principal’s problem, by explicitly identifying the share \( \alpha \) 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 \((F, c) \in A_0 \) and \( \alpha \in [0, 1] \). When \( E_F[y] < c \), the maximum value is 0 (given by \( \alpha = 1 \)). Otherwise, maximizing over \( \alpha \) gives \( \alpha = \sqrt{c/E_F[y]} \), and the objective reduces to

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

Thus, the optimal contract is chosen by taking \((F^*, c^*) \in A_0 \) 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 \( A_0 \) attaining the maximum (a knife-edge case), then there can be several optimal linear contracts.

### 2.4 Discussion of assumptions

Before moving on to the extensions, we should comment on some assumptions, their role in the mechanics of the model and their consequences for interpretation.

The uncertainty on the principal’s part is clearly essential: If the principal knew for certain that \( A = A_0 \), then the optimal contract would in general not be linear (see e.g. Diamond [8]). For example, with \( Y \) finite and \( A \) containing only two actions, the optimal way to incentivize the costlier action would be to pay a positive amount only for the value of output having the highest likelihood ratio, and zero for all other realizations.
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 $A_0$, namely $s_0 = \max_{(F,c) \in A_0} (E_F[y] - c)$, and she could achieve this payoff by setting $w(y) = y - s_0$. However, although there must be some minimum payment to the agent, the assumption that this minimum is 0 is simply a normalization. One could instead assume that the minimum payment is some (positive or negative) constant $w$, and the straightforward analogue of Theorem 2.3 would say that an optimal contract has the form $w(y) = \alpha y + w$. (The nontriviality assumption in this context would say that there exists some $(F, c) \in A_0$ with $E_F[y] - c > w$.)

Likewise, the model assumes that any action must entail a nonnegative cost to the agent. This can be relaxed modestly to allow actions that have private benefits: If we instead assume that the minimum possible cost of an action is $c < 0$, so that an action is defined as an element of $\Delta(Y) \times [c, \infty)$, then the resulting model is equivalent to an instance of the original model with the cost of each action translated by $-c$. Thus, the nontriviality assumption would now require some $(F, c) \in A_0$ with $E_F[y] - c > -c$, and as long as this is satisfied, a linear contract is optimal. (Subsection 3.1 offers another variation on this theme.)

If this version of the nontriviality assumption is not satisfied, then no contract will guarantee the principal any positive payoff. Thus, the model is not suitable (at least immediately) for describing situations where the private benefits from undesirable actions could potentially be very large, e.g. where the agent might be able to steal all the output for personal consumption.

We have also made an assumption of favorable tie-breaking — that if the agent is indifferent among actions, he chooses the best one for the principal. This may seem contrary to the worst-case spirit of the model, but it can be read as a formal shorthand for the standard notion of a contract as consisting of both a payment rule and a recommended action. (Here the recommended action would be contingent on the technology, or we could simply imagine the blanket recommendation “break ties in favor of the principal.”) In any case, other tie-breaking rules would lead to essentially the same results, but may introduce technical complications (e.g. in some instances the optimal contract may not exist, so that $\sup_w V_P(w)$ is approached, but not attained, by linear contracts).

One more, subtler, assumption is hidden in the maxmin expected utility formulation: in contrast to the non-quantifiable uncertainty about the set of possible actions,
for any given action, the risk associated with the action is quantifiable (and moreover, the principal and the agent agree about how to quantify it). Clearly some assumption along these lines is needed: for example, if the agent chose actions by expected utility, but the principal took the worst case over technologies and output realizations, she could never be guaranteed more than 0 payoff. One way to make sense of this combination of non-quantifiable and quantifiable uncertainty is that the risk inherent in any given action depends on physical events occurring in the world, which might be relatively familiar concepts, whereas technologies are too abstract for the principal to be able to reason probabilistically about them. We could also try to appeal to decision-theoretic foundations to justify the maxmin expected utility formulation (see e.g. [24] for references to several such axiomatizations), although such an appeal by itself would not explain why technology appears in the nonprobabilistic parameter space rather than the probabilistic state space.

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: refining the principal’s knowledge by adding a lower bound on the cost of producing any given output distribution, or by otherwise changing the structure of the set of possible technologies; adding a participation constraint; 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 $A$, rather than offering just one contract.

Note that these extensions are independent of each other; we do not pursue the task of writing a single model that is as general as possible.

3.1 Lower bounds on cost

An immediate criticism of the basic model is that it unrealistically allows the agent to produce 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. One way to adjust the model to prohibit this is to suppose instead 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 infimum of $V_P(w|\mathcal{A})$ over all possible technologies $\mathcal{A} \supseteq \mathcal{A}_0$. Everything else is as in the original model. Then, 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 a high level of output deterministically is known to be expensive, but producing the same mean output with high variance might be less costly.) Following Holmström [14], we can also allow there to be other observable variables, besides output, that are informative about the bound on cost. The general result is that the optimal contract is an affine function of output and whatever other relevant variables are observed.

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 separately addressing 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.

### 3.2 Alternative sets of technologies

The basic model assumes that the true technology might be much, much larger than the set of actions known to the principal, since any $\mathcal{A} \supseteq \mathcal{A}_0$ is considered possible. 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 $V_P(w) = \inf_{A \in \mathcal{T}} V_P(w|A)$. Suppose that $\mathcal{T}$ has the following property: For any $A \in \mathcal{T}$, and any arbitrary action $(F, c)$, there exists some $A' \subseteq A$ such that $A' \cup \{(F, c)\} \in \mathcal{T}$. 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

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

and if \( V_P(w) > 0 \), then for any \( F \) attaining the minimum, \( E_F[w(y)] = \inf_{A \in \mathcal{T}} V_A(w|A) \).

(The proof that there exists an optimal contract is slightly more work than before, but one can derive an analogue to (2.3) and check that it is upper semi-continuous in \( \alpha \), which is enough for existence of the optimum.)

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

Although this discussion shows that we need not allow any single agent with a very large technology, it is necessary for the range of actions that might possibly, \( \cup_{A \in \mathcal{T}} A \), to be fairly broad. We cannot (for example) restrict attention to technologies that contain only actions “close” to those in the known technology \( A_0 \) and expect the same results to hold.

### 3.3 Participation constraint

In the basic model, the only constraint that imposed a lower bound on payments to the agent was limited liability. We could instead imagine that there is also a participation constraint, so that the principal is required to guarantee the agent an expected payoff of at least \( \bar{U}_A > 0 \). This could be incorporated by restricting the principal’s maximization problem to contracts \( w \) satisfying \( E_F[w(y)] - c \geq \bar{U}_A \) for some \((F,c) \in A_0\). 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 shows that every contract is weakly outperformed by an affine contract — that is, \( w(y) = \alpha y + \beta \) for some constants \( \alpha, \beta \) with \( \alpha \in [0,1] \) (and strictly so under the full-support condition). 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 \( w' \). However, the step of the original proof in which we change from \( w'(y) = \alpha y + \beta \) to \( \alpha y \) no longer goes through, since the latter contract may not satisfy the constraint.

Let us pursue this further. For any given \( \alpha \), the optimal choice of \( \beta \) is to be as small
as possible subject to the nonnegativity and participation constraints:

$$\beta^*(\alpha) = \max \left\{ 0, \overline{U}_A - \max_{(F,c) \in A_0} (\alpha E_F[y] - c) \right\}.$$  \hfill (3.1)

Evidently, the first case of the max holds when $\alpha$ is greater than some threshold $\overline{\alpha}$, and the second case holds for $\alpha \leq \overline{\alpha}$ (to be precise, $\overline{\alpha} = \min_{(F,c) \in A_0} (\overline{U}_A + c)/E_F[y]$). So we have an analogue of Lemma 2.2: for any $\alpha$, the guarantee of the best affine contract with slope $\alpha$ is at least

$$\max_{(F,c) \in A_0} \left( (1 - \alpha)E_F[y] - \frac{1 - \alpha}{\alpha}c \right) - \beta^*(\alpha),$$  \hfill (3.2)

with equality when the guarantee is sufficiently high. The same arguments as in the proof of Theorem 2.3 now show that the best possible guarantee equals the maximum of (3.2), and is attained by the corresponding affine contract.

But when $\alpha \leq \overline{\alpha}$, (3.2) simplifies to

$$\max_{(F,c) \in A_0} \left( E_F[y] - \frac{1}{\alpha}c \right) - \overline{U}_A,$$  \hfill (3.3)

which is increasing in $\alpha$ (or constant, in the case $c = 0$). And so we conclude that the maximum is attained at some $\alpha \geq \overline{\alpha}$, where $\beta^*(\alpha) = 0$.

Thus, we have an optimal contract that is fully linear, $w(y) = \alpha y$, just as before. Intuitively, as long as the expected payment to the agent in the worst case is pinned down by the binding 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, up to the point where limited liability binds.

### 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 linear separation 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 hope 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

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

and $A^*(w|A)$ is defined as the corresponding argmax.

(ii) The cost is a monetary cost, which the agent directly subtracts from his compensation. Then

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

and $A^*(w|A)$ is the corresponding argmax.

The principal’s payoff under $A$ is defined as $V_P(w|A) = \max_{(F,c) \in A^*(w|A)} E_F[u_P(y - w(y))]$. As before, the principal’s objective is worst-case expected utility, $V_P(w) = \inf_{A \supseteq A_0} V_P(w|A)$.

The nontriviality assumption in specification (i) is that there should exist $(F, c) \in A_0$ with $E_F[u_A(y)] > c$. In specification (ii), we should have $E_F[u_A(y - c)] > 0$. This assumption ensures that it is possible for both parties to have positive expected utility (for example, using a contract $w(y) = \alpha y$ with $\alpha$ close to 1).

We outline the analysis. The zero contract is analyzed as before: If there exists any zero-cost action in $A_0$ then $V_P(0) = \max_{(F,0) \in A_0} E_F[u_P(y)]$, 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

$$V_P(w) = \min F \in \Delta(Y) \text{ such that } E_F[u_A(w(y))] \geq V_A(w|A_0).$$

If $V_P(w) > 0$, then for any $F$ attaining the minimum, $E_F[u_A(w(y))] = V_A(w|A_0)$.

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 $\beta$ 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 $A_0$ 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 $T = \{(u, v) \mid u > V_A(w | A_0), v < V_P(w)\}$ just as before. We obtain $\lambda, \mu > 0$ and $\nu$ 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 $F^\star$. 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

$$\lambda u_A(w'(y)) - \mu u_P(y - w'(y)) = \nu. \quad (3.4)$$

To see this, treat $w'(y)$ as a variable in (3.4). The left-hand side of (3.4) is continuous and strictly increasing; it is $\leq \nu$ when $w'(y) = w(y)$ and tends to $\infty$ 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 $V_P(w' | A) \geq V_P(w)$, so that $V_P(w') \geq V_P$, and the inequality is strict if $A_0$ 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).

Note that it may typically impossible to solve the equation $u_A(w(y)) = \alpha u_P(y - w(y)) + \beta$ explicitly for $w(y)$, even for fairly simple specifications of $U_P$ and $U_A$. However, 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 $y$. (The proofs of these facts are straightforward and are omitted.)

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 $\alpha$) and obtain an improvement. This is no longer possible here, because such a replacement can change the agent’s optimal action under $A_0$ — and therefore the principal’s guarantee from Lemma 3.1 — in unpredictable ways.
3.5 Screening on technology

A natural question for the basic model is: Is the minmax in the principal’s problem equal to the maxmin? That is: we have shown how to find the best possible guarantee for the principal, \( \max_w V_P(w) \); is there actually a single technology \( A \) that pins the principal’s payoff down to this level? If so, this could give an alternative proof of Theorem 2.3. However, the answer turns out to be negative in general:

**Proposition 3.3.** Let \((F^*, c^*) \in A_0\) be the action that maximizes the objective (2.11), and suppose that \( c^* > 0 \). Then there exists \( V_P > \max_w V_P(w) \) such that, for every technology \( A \), there is some contract \( w \) with \( V_P(w|A) \geq V_P \).

**Proof.** We have \( w^*(y) = \alpha^*y \) as the maxmin-optimal contract, with \( \alpha^* = \sqrt{c^*/E_{F^*}[y]} > 0 \), and the principal’s guarantee is \( V_P(w^*) = (\sqrt{E_{F^*}[y]} - \sqrt{c^*})^2 \). Consider any technology \( A \), and let \((F, c)\) be the agent’s action under \( w^* \) and \( A \). Thus

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

(3.5)

We consider two cases.

- If \( c \geq c^*/2 \), then the principal’s payoff from contract \( w^* \) is

\[
(1 - \alpha^*) E_F[y] \geq \frac{1 - \alpha^*}{\alpha^*} (\alpha^* E_{F^*}[y] - c^* + c) = E_{F^*}[y] - 2\sqrt{c^* E_{F^*}[y]} + c^* + \frac{1 - \alpha^*}{\alpha^*} c \\
\geq V_P(w^*) + \frac{1 - \alpha^* c^*}{\alpha^*}.
\]

- Now suppose \( c \leq c^*/2 \). We know that if the principal learns \( A \) before contracting, then by choosing an appropriate contract she can earn at least \((\sqrt{E_F[y]} - \sqrt{c})^2\) (since in fact this is her worst-case guarantee with \( A \) in place of \( A_0 \) — note the condition \( E_F[y] > c \) is met). We show that this expression is bounded strictly above \( V_P(w^*) \). Define

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

(3.6)

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, holding \( c \) fixed, treat \( E_F[y] \) as a variable, constrained
by (3.5) and $E_F[y] > c$. Then $(\sqrt{E_F[y]} - \sqrt{c})^2$ is minimized by taking (3.5) to hold with equality, and in this case $\sqrt{E_F[y]} - \sqrt{c} = g(\sqrt{c})$. Thus we see that the principal can make a payoff of at least

$$\left(\sqrt{E_F[y]} - \sqrt{c}\right)^2 \geq (g(\sqrt{c}))^2 \geq \left(g(\sqrt{c^*/2})\right)^2.$$  

Now observe that $(g(\sqrt{c^*/2}))^2 > (g(\sqrt{c^*}))^2 = V_P(w^*)$.

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

The result of Proposition 3.3 brings to mind the question of screening. That is to say: The timing of the basic model assumes that the principal can only offer a single contract. But if the principal could instead make the contract depend on $\mathcal{A}$ by letting the agent choose among multiple contracts, with different agent types picking different contracts, could she then obtain a strictly higher worst-case payoff guarantee?

It turns out the answer is no, as long as the screening needs to be incentive-compatible. To formalize this, we imagine that the principal offers a menu of contracts $\mathcal{W} = (w_A)$, 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

$$V_A(w_A | \mathcal{A}) \geq V_A(w_{A'} | \mathcal{A}) \quad \text{for all} \quad \mathcal{A}, \mathcal{A'} \supseteq \mathcal{A}_0. \quad (3.7)$$

We write the principal’s worst-case payoff as

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

**Theorem 3.4.** 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 \(V_P(W)\). We may assume that \(A_1 = A_0 \cup \{(F_1, c_1)\}\). Note also that \((F_1, c_1) \notin A_0\), since otherwise \(A_1 = A_0\) and so \(V_P(w_0|A_0) < V_P(W)\) 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 \(A_0\); otherwise he would be willing to take the same action under \(A_1\) as he does under \(A_0\), thereby giving the principal \(V_P(w_0|A_0) \geq V_P(W)\).

Now let \(w_1 = w_{A_1}\), the contract chosen from the menu when the technology is \(A_1\). Under \(w_1\) and \(A_1\), the agent must choose action \((F_1, c_1)\). Proof: If he chooses any action in \(A_0\), then his payoff is at most \(V_A(w_0|A_0)\) (by revealed preference (3.7), since his payoff is the same as \(V_A(w_1|A_0)\)). On the other hand, his payoff under \(w_1\) and \(A_1\) 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 \(A_1\)), which is higher than \(V_A(w_0|A_0)\) by the previous paragraph.

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

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

again by (3.7). Then, the principal’s payoff when the technology is \(A_1\) is

\[
E_{F_1}[y - w_1(y)] = E_{F_1}[y] - c_1 - (E_{F_1}[w_1(y)] - c_1)
\]

\[
\leq E_{F_1}[y] - c_1 - (E_{F_1}[w_0(y)] - c_1)
\]

\[
= E_{F_1}[y - w_0(y)]
\]

\[
< V_P(W)
\]

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

We close this section with a couple additional comments. First, 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.

And second, the gap between the maxmin and the minmax suggests that the principal should be able to improve her worst-case guarantee by deliberately randomizing over contracts. Finding the maxmin randomized contract seems a harder problem. Similarly, it is not clear whether the result of Theorem 3.4 — that screening does not help — persists when the principal can offer menus of randomized contracts. We leave these questions for
future work.

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.

We now return to discuss this model’s potential to help explain the popularity of linear contracts. Many previous scholars have noticed that, whereas theoretical models of agency relations 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 [6, fn. 3] for many references).

One way to try to explain this via our model would be to take the model literally, imagining that contract writers explicitly maximize a worst-case objective, are risk-neutral, and so on. A fuzzier story, but perhaps closer to the truth, is as follows: Just as economists work with stylized models for tractability, so, too, real-life decision-makers may not be able to write down (or solve) their decision problems in full precision. They may therefore be content to adopt a solution that is guaranteed to perform reasonably well in an approximate model (similarly to Simon’s “satisficing” [23]). This paper begins by pointing out how such a guarantee can be obtained from linear contracts, with only slight reliance on knowledge of the environment. Our main result then shows that, while many other contracts can also offer some such guarantee, linear contracts play a distinguished role in this story, by providing the best possible guarantee.

How does our model relate to other explanations for linear contracts in previous literature? The model of Holmström and Milgrom [15] quoted above was one early explanation, also invoking robustness. 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 present this model as capturing robustness, in view of the agent’s large strategy space. However, it is really the stationary 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 [8] gives an argument particularly close to the intuition of this paper. Di-
amond’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 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 restrictive. Furthermore, 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.

Several other papers consider models where the contractible outcome variable combines effort with mean-zero additive noise, leading naturally to linear contracts. For example, a version of the model of Edmans and Gabaix [9] with linear utility and additive noise gives this result. However, their model focuses mainly on implementing a particular action, rather than the more primitive objective of maximizing the principal’s payoff. Earlier works by Laffont and Tirole [17] and McAfee and McMillan [20] 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 again an optimal menu in which payment is linear in output within each contract. Again, however, there may also be other optimal menus. In any case, the assumption of additive noise is quite specific.

Chassang [4] considers a dynamic environment with risk-neutrality and limited liability, and gives the same lower bound (2.1) on the performance of a linear contract, by the same calculation as given here. Chassang also gives a maxmin optimality result for linear contracts in a certain class of environments (his Corollary 1); in that class, first-best total surplus may be arbitrarily small, so the objective used is the ratio of the principal’s profit to first-best surplus. The argument given for Chassang’s optimality result is essentially by construction of a parameterized environment in which one directly calculates an upper bound on the principal’s payoff, in contrast to the argument for Theorem 2.3 here, which follows the economic intuition about the alignment between the principal’s and agent’s payoffs.

Finally, Hurwicz and Shapiro [16] also consider a maxmin contracting problem whose objective involves the ratio of principal’s profits to first-best total surplus. They focus on a particular class of environments involving quadratic effort costs. Their paper does not discuss economic intuition behind the optimality argument, which involves a differential
inequality; it seems quite different from the argument here.

Against this backdrop, then, the contribution of the current paper is a specific combination of features: The model allows many degrees of freedom (the set of known actions the agent has can be arbitrary, with no functional form assumptions); the concern for robust performance is modeled explicitly through the maxmin payoff objective; and we give a mathematical argument for optimality that reflects the simple intuition that linearity works by aligning the agent’s goal with the principal’s. Also, in contrast to previous maxmin results, the simple expected-profit objective here might also be considered more natural than the ratio objective, although this distinction is a matter of taste.

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 same modeling approach to the principal-expert problem of Zermeño [26, 27] to study worst-case-optimal incentives for information acquisition.

Relatedly, the modeling approach here may prove useful to future 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. However, this suggestion should be supplemented with a note of caution: It is common in applied theory models to assume full knowledge of the environment, but then exogenously impose a restriction to linear contracts (e.g. [10, 13]). The model here cannot be invoked as a microfoundation for this practice, since the contract that is best among all linear contracts when the technology is known to be $A_0$ is generally different from the maxmin-optimal contract studied here.

A An alternative approach

We give here another, more direct 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 $A_0$. 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

$$E_{F_0}[\alpha y] - c_0 = E_{F_0}[w(y)] - c_0 = V_A(w | A_0),$$
and the principal then earns

$$E_{F_0}[(1-\alpha)y] = E_{F_0}[y - w(y)] = V_P(w|A_0) \geq V_P(w).$$

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

$$(1-\alpha)E_F[y] \geq (1-\alpha)E_{F_0}[y] = V_P(w|A_0) \geq V_P(w).$$

Also, we have \( E_F[w'(y)] - c \geq V_A(w'|A_0) \geq V_A(w|A_0) \) by optimality for the agent; and if equality holds throughout, then the agent would also be willing to choose \((F_0, c_0)\), again giving the principal at least \( V_P(w) \); thus \( V_P(w'|A) \geq V_P(w) \) in this case too. So we can focus on the case when \( E_F[y] < E_{F_0}[y] \) and \( E_F[w'(y)] - c > V_A(w|A_0) \).

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 \( A_0 \cup \{(F', c)\} \). The agent’s payoff from \((F', c)\) is

$$E_{F'}[w(y)] - c = \lambda E_{F_0}[w(y)] + (1-\lambda)w(0) - c$$

$$\geq \lambda E_{F_0}[w(y)] - c$$

$$= \lambda \alpha E_{F_0}[y] - c$$

$$= \alpha E_F[y] - c$$

$$= E_F[w'(y)] - c$$

$$> V_A(w|A_0)$$

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

$$E_{F'}[y - w(y)] = \lambda E_{F_0}[y - w(y)] - (1-\lambda)w(0)$$

$$\leq \lambda E_{F_0}[y - w(y)]$$

$$= (1-\alpha)E_F[y]$$

$$= E_F[y - w'(y)]$$

$$= V_P(w'|A).$$

Thus we have \( V_P(w) \leq V_P(w'|A) \). So we have shown this inequality holds for all \( A \),
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 $\delta_0$ 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 $F \in \Delta(Z)$ and an associated cost $c$. 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 $\min\{z_1 \mid z \in Z\} = 0$. (Let $\bar{z}$ denote some fixed element of $Z$ with $\bar{z}_1 = 0$.) Also, note that we can use this same framework to model more complex restrictions on costs when there are less than $k$ degrees of freedom in the observable variables. 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 putting

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

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 $(F, c) \in \Delta(Z) \times \mathbb{R}^+$ satisfying $c \geq b(E_F[z])$. A technology is a compact set of actions. A technology $A_0$, 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 be-
fore. Given contract $w$ and technology $A$, the agent’s utility is $V_A(w|A) = \max_{(F,c) \in A} (E_F[w(z)] - c)$ and his choice set is $A^*(w|A) = \arg\max_{(F,c) \in A} (E_F[w(z)] - c)$. The principal’s expected payoff under $A$ is $V_P(w|A) = \max_{(F,c) \in A} (E_F[w(z)] - c)$. The principal’s objective, $V_P(w)$, is then defined to be the infimum of $V_P(w|A)$ over all technologies $A \supseteq A_0$.

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 $\alpha_i$ and $\beta$.

In the setting in Subsection 3.1, 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 requires us to deal carefully with the boundary case in which the agent’s best action under any possible technology is already available in $A_0$. Previously, this case mattered only for the zero contract, 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 in $F$.

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

(i) $V_P(w) = \min_{F \in \Delta(Z)} E_F[z_1 - w(z)]$ over $F \in \Delta(Z)$ such that $h(F|w) \geq V_A(w|A_0)$.

(ii) $\max_{F \in \Delta(Z)} h(F|w) = V_A(w|A_0)$.

**Proof.** Let $F_0$ be a distribution attaining the minimum in (i). (The constraint set is nonempty since it is satisfied by the action chosen under $A_0$.) 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 $\epsilon$. By concavity, $h(F'|w) \geq (1 - \epsilon)h(F_0|w) + \epsilon h(F_1|w) > h(F_0|w)$. So if $A = A_0 \cup \{(F', b(E_{F'}[z]))\}$, then the agent’s unique optimal action in $A$ 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 (i). On the other hand the principal is guaranteed at least this much, just as in the proof of Lemma 2.1.

Also, if \( h(F_0|w) > V_A(w|A_0) \) strictly, then let \( A = A_0 \cup \{ (F_0, b(E_{F_0}[z])) \} \). 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 (i). Thus in either of these situations \( V_P(w) \) is as specified by conclusion (i).

We are left with the situation in which \( F_0 \) maximizes \( h(F|w) \) over all \( F \in \Delta(Z) \) and \( h(F_0|w) = V_A(w|A_0) \). In this case, we have 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

\[
t(z) = \max\{b(z) + V_A(w|A_0), z_1 - V_P(w)\}
\]

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 \((z, w(z))\), for \( z \in Z \). 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 \( E_F[w(z)] > t(E_F[z]) \). In particular,

\[
E_F[w(z)] > b(E_F[z]) + V_A(w|A_0)
\]

implying

\[
h(F|w) > V_A(w|A_0),
\]
and also
\[ E_F[w(z)] > E_F[z_1] - V_P(w) \]

implying
\[ E_F[z_1 - w(z)] < V_P(w). \]

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
\[
\sum_i \lambda_i z_i + \mu c \leq \nu \quad \text{for all} \quad (z, c) \in S, \tag{B.1}
\]
\[
\sum_i \lambda_i z_i + \mu c \geq \nu \quad \text{for all} \quad (z, c) \in T, \tag{B.2}
\]

and some \( \lambda_i \) or \( \mu \) is nonzero. Inequality (B.2) 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.1) gives \( \sum_i \lambda_i z_i \leq \nu \) for all \( z \in Z \), while (B.2) 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 \( \lambda_i \) are zero, this contradicts the full-dimensionality of \( Z \).

Now we can rewrite (B.1) as
\[
w(z) \leq \frac{\nu - \sum_i \lambda_i z_i}{\mu} \quad \text{for all} \quad z \in Z.
\]

This motivates us to define
\[
w'(z) = \frac{\nu - \sum_i \lambda_i z_i}{\mu},
\]
an affine contract satisfying \( w' \geq w \) pointwise.

Now we are ready to check that \( V_P(w') \geq V_P(w) \). Let \((F_0, c_0)\) be the action that the agent chooses under \( w \) and technology \( A_0 \).

Consider any technology \( A \supseteq A_0 \). As in the original proof of Theorem 2.3, we certainly have have \( V_A(w'|A) \geq V_A(w'|A_0) \geq V_A(w|A_0) \). Let \((F, c)\) be the action chosen under \( w' \)
and \(A\). Then (B.2) implies

\[
t(E_F[z]) \geq \frac{\nu - \sum_i \lambda_i E_F[z_i]}{\mu} = E_F[w'(z)] = V_A(w'|A) + c \geq V_A(w|A_0) + c \geq b(E_F[z]) + V_A(w|A_0).
\]

If the inequality is strict, then \(t(E_F[z]) = E_F[z_1] - V_P(w)\), and so we have

\[
V_P(w'|A) = E_F[z_1 - w'(z)] = t(E_F[z]) + V_P(w) - E_F[w'(z)] \geq V_P(w).
\]

Otherwise, \(t(E_F[z]) = b(E_F[z]) + V_A(w|A_0)\) and so all the inequalities in the stacked chain above are equalities. In particular, the second inequality is an equality, implying \(V_A(w'|A) = V_A(w'|A_0) = V_A(w|A_0)\). Since the agent does at least as well as \(V_A(w|A_0)\) by taking action \((F_0, c_0)\), this action is in his choice set under \(w'\) and \(A\), and so the principal gets at least the corresponding payoff: \(V_P(w'|A) \geq E_{F_0}[z_1 - w'(z)]\). This is equal to \(E_{F_0}[z_1 - w(z)]\), since otherwise \(V_A(w'|A_0) > V_A(w|A_0)\). But \(E_{F_0}[z_1 - w(z)] = V_P(w|A_0) \geq V_P(w)\).

Thus in either case, \(V_P(w'|A) \geq V_P(w)\). This holds for all \(A\), so \(V_P(w') \geq V_P(w)\).

**Case (ii).** In this case, define \(S\) to be the convex hull of all pairs \((z, w(z))\), and \(T\) to be the set of all \((z, c)\) with \(z\) in the convex hull of \(Z\) and \(c > b(z) + V_A(w|A_0)\). These are convex, and disjoint: otherwise, there exists \(F\) such that

\[
E_F[w(z)] > b(E_F[z]) + V_A(w|A_0)
\]

which reduces to

\[
h(F|w) > V_A(w|A_0),
\]

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

\[
\sum_i \lambda_i z_i + \mu c \leq \nu \quad \text{for all} \quad (z, c) \in S,
\]

\[
\sum_i \lambda_i z_i + \mu c \geq \nu \quad \text{for all} \quad (z, c) \in T,
\]

(B.3) (B.4)
and we show that $\mu > 0$. Again, (B.3) implies

$$w(z) \leq \frac{\nu - \sum_i \lambda_i z_i}{\mu}$$

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

$$E_F[w'(z)] - c \leq E_F[w'(z)] - b(E_F[z]) = w'(E_F[z]) - b(E_F[z]) \leq V_A(w|A_0)$$

where the second inequality follows from (B.4). That is, the agent can never earn a higher expected payoff than $V_A(w|A_0)$. On the other hand, the agent can always earn at least this much, since $V_A(w'|A) \geq V_A(w'|A_0) \geq V_A(w|A_0)$ as usual. So we have equality. From here the argument finishes just as at the end of case (i), and we have $V_P(w') \geq V_P(w)$.

**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 $\bar{b} = \max_{z \in Z} b(z)$ and $\bar{y} = \max(Y)$. Note that for any contract $w$ satisfying $\max_{z \in Z} w(z) - \bar{b} \geq \bar{y}$, the agent can potentially attain a payoff greater than $\bar{y}$, which means that the principal cannot be guaranteed a positive payoff. Hence we can restrict attention to contracts with $w(z) \in [0, \bar{y} + \bar{b}]$ for all $z$. By full-dimensionality, this implies a compact range of possible values for $\alpha$ and $\beta$. 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 $\alpha, \beta$ under this topology, it will then follow that $V_P(w)$ is upper semi-continuous in $\alpha, \beta$, so that the maximum is attained.

Let $w_1, w_2, \ldots$ be any contracts that converge to some contract $w_\infty$ in the sup norm. We wish to show that $V_P(w_\infty) \geq \lim \sup_k V_P(w_k)$. 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 $A$, and let $(F_k, c_k)$ be the agent’s chosen action under $A$ and contract $w_k$. We may again pass to a subsequence and assume that $(F_k, c_k)$ has some limit $(F_\infty, c_\infty) \in A$. Then straightforward continuity arguments show that $(F_\infty, c_\infty)$ is an optimal action (perhaps not the only one).
for the agent under \( w_\infty \), and its payoff to the principal is the limit of the corresponding payoffs of \((F_k, c_k)\) under \( w_k \). Hence,

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

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 (3.4) is a contract, we need to check that it is continuous. In fact we show more. For every real number \( y \), and all \( \lambda > 0, \mu \geq 0, \nu \in \mathbb{R} \), 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 \( \lambda, \mu, \nu \). Indeed: within any compact region of \((y, \lambda, \mu, \nu)\)-space, (C.1) implies \( w \leq w'(y) \leq \bar{w} \) for some bounds \( w, \bar{w} \). 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 \( \tilde{w}' \neq w'(y; \lambda, \mu, \nu) \). Then, by continuity, (C.1) holds at \((y, \lambda, \mu, \nu)\) for both \( w'(y; \lambda, \mu, \nu) \) and \( \tilde{w}' \), 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 \((\alpha, \beta)\). (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 makes the principal pay more than $C$ and thereby forces her 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 $(\alpha, \beta)$. The principal’s guarantee, $V_P(w)$, is in turn upper semi-continuous in $w$ under the sup-norm topology on contracts (this is the same argument as at 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 $w$ varies continuously (under this same topology) with respect to the parameters $\alpha, \beta$. Therefore, the optimum exists.

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 maximin-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 $\mathcal{A}$.

First 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 (3.6); 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

$$\sqrt{E_{F}[y]} - \sqrt{c} > g(\sqrt{c}) \geq \sqrt{E_{F^*}[y]} - \sqrt{c^*}$$

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 \( E_F[y] \leq (1 - \alpha^*)E_{F^*}[y] \leq (1 - \alpha^*)\bar{y} \).

- Otherwise, let \( F' \) be a mixture of \( F \) and \( \delta_0 \), with weight \( (1 - \alpha^*)E_{F^*}[y] / E_F[y] \) on \( F \) and the remaining weight on \( \delta_0 \). We claim that \((F', 0) \in \mathcal{A}\). Indeed:
  
  \[
  \alpha^*E_{F'}[y] - 0 = \alpha^*(1 - \alpha^*)E_{F^*}[y]E_F[y] = \alpha^*E_{F^*}[y] - \alpha^{*2}E_{F^*}[y] = \alpha^*E_{F^*}[y] - c^*.
  \]

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

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

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 \( \alpha^* \) when we recall \( c^* = \alpha^{*2}E_{F^*}[y] \)). Therefore, the principal’s payoff is

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

This shows that under technology \( \mathcal{A} \), no other contract can do better than \( w^* \) for the principal, as claimed.

References


