Robust Contracting with Additive Noise

Gabriel Carroll Delong Meng
Stanford University

October 1, 2016

Abstract

We investigate the idea that linear contracts are reliable because they give the same incentives for effort at every point along the contract. We ask whether this reliability leads to a microfoundation for linear contracts, when the principal is profit-maximizing. We consider a principal-agent model with risk neutrality and limited liability, in which the agent observes the realization of a mean-zero shock to output before choosing how much effort to exert. We show that such a model can indeed provide a foundation for reliable contracts, and illustrate what elements are required. In particular, we must assume that the principal knows a lower bound, but not an upper bound, on the shocks.

Keywords: linear contract, principal-agent problem, robustness, worst-case
JEL Classification: D81, D82, D86

Authors are listed in random order. This project developed from conversations with Tomasz Sadzik. We also thank (in random order) Alessandro Pavan, Zhengyu Huang, Alex Wolitzky, and anonymous referees for helpful comments.

1 Introduction

Why are certain forms of incentive contracts particularly widespread in the world? A significant strand of literature in agency theory has attempted, in particular, to understand why *linear* contracts — in which the agent is paid some fixed fraction of the output he produces — are common, where textbook models tend to predict contract structures that

are more finely tuned to the environment. (See e.g. Bhattacharyya and Lafontaine (1995, pp. 763–4) and Chu and Sappington (2007, fn. 3) for examples of the prevalence of linear contracts.) Much of this literature has emphasized robustness to uncertainty or to large spaces of possible actions by the agent as a way to microfound linear contracts (Hurwicz and Shapiro, 1978; Holmström and Milgrom, 1987; Chassang, 2013; Carroll, 2015).

In this paper we explore the possibility of foundations based on the following intuitive property of linear contracts: the marginal incentive for effort is the same no matter where along the contract the agent currently stands. That is, we consider a model in which the agent privately observes some shock ϵ that affects output; the agent then chooses effort x at a cost $c(x)$, and the principal observes total output $x + \epsilon$. In this environment, linear contracts have the property that they incentivize the same effort choice no matter what ϵ has been observed; we call this property *reliability*. Reliability is convenient since it makes the contract easy to analyze. But of course, the principal's goal is not analytical convenience, but rather profit. So, we ask whether reliability can provide a justification for using linear contracts, when the principal is ultimately concerned with profit.

One can imagine numerous situations in which it is reasonable that the agent observes a shock before choosing effort. For example, the agent may be part of a team, and observe the effort of his teammates, but the principal only sees total output. Or output may be measured in terms of the time for a project to reach completion, and the agent may observe exogenous delays that are going to occur (as in the highway procurement studied in Lewis and Bajari (2014)). Another application is to investor-entrepreneur contracts. The investor cannot observe the firm's total cash flow, but the entrepreneur can, and the entrepreneur decides how much fund to divert to private benefits (see DeMarzo and Fishman, 2007; Biais et al., 2007). Other situations such as sharecropping and trading voyages are discussed in Lacker and Weinberg (1989).

What are our findings? A profit-motivated principal may indeed choose a reliable contract, under appropriate circumstances. We avoid income effects; our model assumes risk-neutrality (and limited liability).[1] And we also assume that the principal does not even know the distribution of the additive shock ϵ, but knows only that it has mean zero; the principal maximizes her worst-case expected profit, over all such distributions. In such a model, linear contracts are a natural candidate for the maxmin optimum because they perfectly hedge the principal's uncertainty: Since they always induce the same effort

[1] A model with an interim participation constraint would deliver very similar results. (As it is, our model could be interpreted as having an ex-ante participation constraint, where the outside option is low enough to be non-binding.)

choice, they always lead to the same expected output, and the same expected payment to the agent, regardless of the noise distribution; and so they always give the same expected net profit for the principal. This intuition naturally connects linear contracts' reliability to their payoff for the principal. In such a model, we show that reliable contracts are indeed optimal — and we identify the worst-case noise distribution.

Even so, the result relies on a combination of assumptions, as we discuss further in the conclusion. In particular, we must build an asymmetry into the model: we assume that the noise is bounded below and not above. If noise is bounded above also, then reliable contracts are suboptimal; we characterize the optimum. If noise is unbounded on both sides, the principal cannot be guaranteed any positive profit. Moreover, the optimal contract, even when reliable, is not linear; a piecewise-linear contract can do better.

Numerous previous models have exploited the reliability feature of linear contracts. For example, Laffont and Tirole (1986) and McAfee and McMillan (1987) consider problems in which the agent's effort is perturbed by additive noise, and the optimal contract is linear in observed output. However, providing incentives is not costly in their models in the same way that it is in ours: we assume a limited liability constraint; they do not, so the only binding constraint imposing a lower bound on payments in their models is an ex-ante participation constraint. In our model, getting rid of limited liability in this way would lead to the trivial solution of selling the firm to the agent. (In Laffont and Tirole (1986) and McAfee and McMillan (1987), there is also a screening component that makes the models more complex.) The classic model of Holmström and Milgrom (1987), in which the agent controls the drift of a Brownian motion that is fully observed, also shares the reliability intuition — linear contracts are optimal because the incentives at each point in time are independent of the history; the counterpart of our shock ϵ is the realized path of the output so far. Their model also assumes no limited liability; they have risk aversion, which leads to a different tradeoff than limited liability does in ours.

The most closely connected work in this line is Edmans and Gabaix (2011). Their model, in which the agent observes the shock before choosing effort, is the basis for ours, although they allow for risk-aversion and multiple periods, so that our model is a very special case. Edmans and Gabaix take as exogenously given the effort level that the principal wishes to implement. In general, this effort level may be a function of the shock realization, but their main model assumes that the target effort is independent of the shock. Our exploration here is concerned with possible microfoundations for this assumption.

Several other recent papers on dynamic contracts also study settings where an agent

could privately observe a signal before choosing his effort. Edmans et al. (2012) extend the model of Edmans and Gabaix (2011) to allow for private savings in CEO compensation. They assume that the principal implements constant effort in each period, and they find that in the optimal contract the agent's consumption is log-linear in terms of the firm's return. Again, our work serves as a potential micro-foundation for this "constant effort" assumption. Garrett and Pavan (2012, 2015) study seniority-based pay and managerial turnovers, using models where an agent privately observes his productivity before choosing the effort. Their papers employ the dynamic nature of the contracting environment, and effort wouldn't be constant over time. In Garrett and Pavan (2012) the optimal contract awards longer-tenure managers more high power schemes, and optimal effort on average could change over time.

There is also a closely connected strand of the cost-based procurement literature following Laffont and Tirole (1986). In particular, Rogerson (2003) and Chu and Sappington (2007) consider a simplified version of the Laffont-Tirole model that is mathematically very similar to our model, but with a known distribution of ϵ (which is interpreted there as the agent's innate productivity). They show that for some salient specifications of the model, piecewise-linear contracts with two segments are approximately optimal. Garrett (2014) shows that such contracts are actually optimal in a maxmin version of the model with uncertainty about the agent's cost function. The contracts derived in these papers, when translated back into our setting, are piecewise-linear contracts whose initial segment is flat — just like the reliable contracts featured in this paper. However, in the contracts studied in these papers, some agent types exert high effort and others no effort. In our piecewise-linear contracts, the agent always exerts the same level of effort, and the flat portion exists only to satisfy limited liability (and indeed, piecewise-linearity here is only one choice among many equally good ways to satisfy this constraint).

Finally, our paper fits in the broader literature on mechanism design with maxmin objectives, starting with López-Cuñat (2000) and followed by many others, e.g. Chung and Ely (2007); Frankel (2014); Garrett (2014). Many of these papers show how, in various settings, an intuitively simple mechanism can emerge as a solution to a problem where a principal wants a mechanism that is robust to some uncertainty about the environment, expressed via a maxmin objective. Previous work by one of the authors (Carroll, 2015) has given maxmin foundations for linear contracts, using quite different ideas from this paper: There, potential actions by the agent correspond to arbitrary distributions over output (in contrast to the one-dimensional effort choice and additive functional form assumed here), and the principal faces large-scale uncertainty about the space of actions that are actually

available to the agent. Linear contracts are robust in that setting because they tightly tie the agent's expected payoff to the principal's, regardless of what shape of distribution the agent chooses. In effect, the key linear relation there is between the principal's payoff and the agent's, whereas here it is between the principal's payoff and the exogenous ϵ.

2 The model

The environment is given by the following parameters: bounds $a, b \in \mathbb{R} \cup \{\pm\infty\}$ on the possible shocks, where $a < 0 < b$; a maximum effort level $\overline{x} > 0$; and a cost-of-effort function $c : [0, \overline{x}) \to \mathbb{R}^+$. We assume c is increasing, strictly convex, twice-differentiable, $c(0) = c'(0) = 0$, and $c(x) \to \infty$ as $x \to \overline{x}$. All these parameters are assumed to be common knowledge between our two characters, the principal and agent.

The agent will privately observe a shock $\epsilon \in [a, b]$, and then choose an effort level x at cost $c(x)$. The publicly observed output is $y = x + \epsilon$, and this is what may be contracted on. Both parties are risk-neutral, and limited liability applies: the agent can never be paid less than zero. (We do not explicitly include a participation constraint; see footnote 1.) Output will always lie in $[a, b + \overline{x}]$; thus, a *contract* is an upper semi-continuous function $w : [a, b + \overline{x}] \to \mathbb{R}^+$. The upper semi-continuity requirement is imposed to avoid problems of best-reply nonexistence.

When the agent observes shock ϵ, the agent chooses effort x to maximize his net payoff $w(x + \epsilon) - c(x)$. We write $\mathbf{x}(w; \epsilon)$ for the set of such maximizers x (there may be more than one). We then write

$$u_P(w; \epsilon) = \max_{x \in \mathbf{x}(w;\epsilon)} (x + \epsilon - w(x + \epsilon))$$

for the principal's resulting payoff. Note that the agent may have several optimal choices of effort, in which case we assume he chooses in a way that is best for the principal; this is what the max formulation represents.

Finally, the principal is uncertain about the distribution of ϵ. All she knows is its mean, which (by normalization) we can take to be zero. Thus, write $\mathcal{F}[a, b]$ for the set of mean-zero distributions on $[a, b]$. The principal wants to choose a contract that guarantees her a reasonable payoff in expectation, in spite of her uncertainty. In particular, she evaluates any contract in terms of its worst-case guarantee:

$$V_P(w) = \inf_{F \in \mathcal{F}[a,b]} (\mathbb{E}_F[u_P(w; \epsilon)]) .$$

One possible choice for a contract is a *linear* contract, which in this case would take the form $w(y) = \alpha(y - a)$. Under such a contract, the agent would always choose effort $x^*(\alpha)$ given by the first-order condition $c'(x) = \alpha$. Consequently, we have $u_P(w; \epsilon) = (1 - \alpha)(x^*(\alpha) + \epsilon) + \alpha a$. When we take the average with respect to *any* distribution $F \in \mathcal{F}[a, b]$, the ϵ term drops out, and consequently we have $V_P(w) = (1 - \alpha)x^*(\alpha) + \alpha a$.

More generally, let us say that a contract w *reliably implements* effort level x^* if $x^* \in \mathbf{x}(w; \epsilon)$ for all ϵ; and call a contract *reliable* if it reliably implements some x^*. Thus, linear contracts are reliable. However, a linear contract is not the only way — and more importantly, not the cheapest way — to implement a given effort level. With a linear contract, even at the lowest shock realization $\epsilon = a$, the agent earns a positive rent, $\alpha x^* - c(x^*)$. By paying less for output levels below $a + x^*$ (which the agent should never produce anyway), the principal can lower this rent.

Lemma 2.1. *A contract w reliably implements effort $x^* > 0$ if and only if there is some constant $h \geq 0$ such that the following hold:*

- $w(y) \leq c(y - a) + h$ *for all* $y \in [a, a + x^*]$;
- $w(y) = c'(x^*)(y - a) + h + c(x^*) - c'(x^*)x^*$ *for all* $y \in [a + x^*, b + x^*]$;
- $w(y) \leq c(y - b) + c'(x^*)(b - a) + h$ *for all* $y \in [b + x^*, b + \overline{x})$.

In this case,

$$V_P(w) = x^* + c'(x^*)a - c(x^*) - h.$$

The (straightforward) proof is in Appendix A. Note that the lemma excludes the case $x^* = 0$; in this case, any contract w that's non-increasing can reliably implement x^*.

In particular, the optimal way for the principal to reliably implement any given effort x^* is by setting $h = 0$. For example, the piecewise-linear contract given by

$$w(y) = \max\{c'(x^*)(y - a) + c(x^*) - c'(x^*)x^*, 0\}$$

accomplishes this. Such a piecewise-linear contract is depicted in Figure 1. The thin line $\widehat{w}(y)$ is a linear contract, and the thick lines $w(y)$ represent a piecewise-linear contract. As shown, the piecewise-linear contract makes a non-local incentive constraint bind: at the lowest shock realization $\epsilon = a$, the agent is indifferent between effort x^* and 0, unlike in a linear contract.

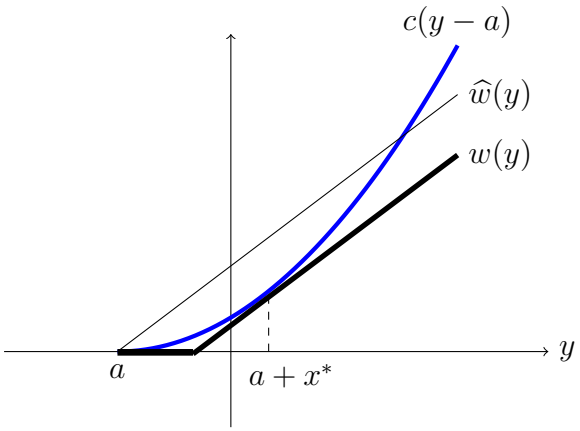

Figure 1: Linear vs. piecewise-linear contracts

Note also from the formula for $V_P(w)$ that if a is close enough to zero, the principal's problem is nontrivial — that is, she can indeed get a positive profit guarantee by using some such reliable contract. (In our normalization, zero is not the principal's outside option but rather the payoff that the she would get by always paying the agent the minimum possible and inducing no effort. So "positive profit" means it is profitable to induce some effort.)

However, our principal is not specifically interested in reliability; she is interested in maximizing her profit guarantee $V_P(w)$. Can this lead her to choose a reliable contract? Our main result is an affirmative answer — *if* we assume that b, the upper bound on shocks, is infinite.

Theorem 2.2. *Suppose that $b = \infty$. Then, the maximum value of $V_P(w)$ is attained by a reliable contract.*

The profit guarantee is equal to $\max_x (x - c(x) + c'(x) \cdot a)$. It is positive if a is close enough to 0; it decreases as a decreases, and it goes to 0 as a goes to $-\infty$.

We prove this theorem momentarily.

As a side point, we observe that as $a \to 0$, the optimal profit guarantee approaches the first-best surplus. This should not be surprising, since $a = 0$ forces $\epsilon = 0$ for certain and thus represents a complete-information model.

3 The main proof

So far we have described contracts with w as a function of output y. This expository approach is in keeping with prior literature exploiting the reliability property (e.g. Holmström and Milgrom, 1987; Edmans and Gabaix, 2011), and makes the structure of linear contracts immediately apparent. However, the economics of our problem is naturally that of a screening model: the agent's type is his private information ϵ, and he chooses an "allocation" y to produce and a corresponding payment $w(y)$ to receive. Thus, in the screening formulation of the problem, a contract would be written as a menu of pairs $(y(\epsilon), w(y(\epsilon)))$, such that it is incentive-compatible for each type ϵ to choose the corresponding pair.

The "value" of choosing y for an agent of type ϵ is $-c(y-\epsilon)$ (or $-\infty$ if $y \notin [\epsilon, \epsilon+\overline{x})$). The standard single-crossing condition then requires that this function should have increasing differences in y and ϵ. This is assured by the convexity of c. Consequently, standard arguments imply that, for any incentive-compatible contract, $y(\epsilon)$ must be weakly increasing; and conversely, any allocation rule $y(\epsilon)$ that is weakly increasing and satisfies $y(\epsilon) \in [\epsilon, \epsilon+\overline{x})$ can be supported by the transfers given by the standard integral formula arising from the envelope theorem.

Let $U(\epsilon) = w(y(\epsilon)) - c(y(\epsilon)-\epsilon)$ denote the agent's utility when his type is ϵ. The standard envelope integral formula gives

$$U(\epsilon) = U(a) + \int_a^{\epsilon} c'(y(s)-s) \cdot ds.$$

Consequently the transfer rule is given by

$$w(y(\epsilon)) = c(y(\epsilon)-\epsilon) + U(a) + \int_a^{\epsilon} c'(y(s)-s) \cdot ds. \tag{3.1}$$

Before we proceed, we must address a technical issue involving (3.1): Although it is true that any increasing allocation rule $y(\epsilon)$ can be implemented by transfers (3.1), in our setting this may not be the *unique* such transfer rule for a given $y(\epsilon)$. Specifically, at values where $y(\epsilon) - \epsilon = 0$, it is impossible for higher types $\epsilon' > \epsilon$ to imitate type ϵ, because the effort $y(\epsilon) - \epsilon'$ would need to be non-negative. In that case, the standard logic to pin down U breaks down because even if we assume U is differentiable, we only know that $U'(\epsilon) \leq c'(y(\epsilon)-\epsilon)$, without necessarily having equality. Then, (3.1) is not unique in implementing $y(\epsilon)$. (To see this clearly, recall that the schedule of zero effort for all ϵ is implemented by any transfer rule that is decreasing.)

This means we cannot immediately rewrite the principal's problem in terms of the allocation rule $y(\epsilon)$ only. However, we can overcome this issue by focusing on the optimal contract. Lemma C.1 shows that if a contract performs better than any reliable contract, then it must induce effort bounded away from zero, for all ϵ. Hence either the optimal contract is reliable, in which case (3.1) holds by Lemma 2.1; or we have $\inf_\epsilon(y(\epsilon)-\epsilon) > 0$, and (3.1) again becomes an equality.[2] Therefore for the optimal contract, we know that Equation (3.1) holds.

We now formulate the principal's problem. The optimal contract guarantees the principal a profit equal to

$$\inf_{F\in\mathcal{F}[a,b]} \mathbb{E}_F[y(\epsilon) - w(y(\epsilon))] = \inf_{F\in\mathcal{F}[a,b]} \mathbb{E}_F\left[y(\epsilon) - c(y(\epsilon)-\epsilon) - U(a) - \int_a^\epsilon c'(y(s)-s)\cdot ds\right].$$

In the optimal contract we set $U(a) = 0$. For the ease of notation, define $x(\epsilon) = y(\epsilon) - \epsilon$. If F has a density function f, then the expectation can be written in terms of virtual values, as follows:

$$\int_a^b \left(x(\epsilon) - c(x(\epsilon)) - \frac{1-F(\epsilon)}{f(\epsilon)}\, c'(x(\epsilon))\right)\cdot dF(\epsilon). \tag{3.2}$$

Theorem 2.2 claims that when $b = +\infty$ the maxmin value of expression (3.2) is given by a reliable contract, and the maxmin profit is equal to $\max_x (x - c(x) + c'(x)\cdot a)$.

The proof idea, given the background developed at this point, is simple: We know that a reliable contract gives the same expected profit no matter what distribution F the principal faces. To show that no other contract gives a better contract guarantee, it suffices to find a particular F such that the reliable contract is optimal for that F (i.e. to show the principal's problem has a saddle point). From looking at (3.2), we see that if F is chosen so that the virtual value is the same for every ϵ, then indeed the optimum will be to have $x(\epsilon)$ constant; thus we conjecture an exponential distribution for our worst-case F. Then all that remains is to check that the details work out.

[2]Here the argument is standard once again: Assume $\inf_\epsilon(y(\epsilon)-\epsilon) > 0$. We know y is increasing. Also, on any interval, $y(\epsilon)-\epsilon$ must be bounded strictly below $\overline{x}$, since inducing effort close to $\overline{x}$ would require paying an arbitrarily large amount and so lead to arbitrarily negative profits. Now, for any ϵ_0, choose a small enough neighborhood $\mathcal{N}$ of ϵ_0 and a small positive number $\eta > 0$ such that $\eta < y(\epsilon)-\epsilon' < \overline{x}-\eta$ for all $\epsilon, \epsilon' \in \mathcal{N}$. Then IC constraints imply that $c(y(\epsilon)-\epsilon)-c(y(\epsilon)-\epsilon') \le U(\epsilon')-U(\epsilon) \le c(y(\epsilon')-\epsilon)-c(y(\epsilon')-\epsilon')$, which is the standard local constraint for the integral envelope formula. This implies that U is Lipschitz-continuous on $\mathcal{N}$, with Lipschitz constant $c'(\overline{x}-\eta)$. Hence U is absolutely continuous, with derivative equal to $U'(\epsilon) = c'(y(\epsilon)-\epsilon)$ wherever it exists. We obtain that $U(\epsilon) = U(a) + \int_a^\epsilon c'(y(s)-s)\cdot ds$.

Proof of Theorem 2.2. We first construct a reliable contract that guarantees profit $\max_x (x - c(x) + c'(x) \cdot a)$. Let x^* denote a maximizer of $x - c(x) + c'(x) \cdot a$. Let w^* denote a reliable contract that implements x^*. Lemma 2.1 implies that

$$\inf_{F \in \mathcal{F}[a,b]} \mathbb{E}_F[u_P(w^*; \epsilon)] = x^* - c(x^*) + c'(x^*) \cdot a.$$

We next show that no contract can guarantee a larger profit. All we need to do is to construct a distribution F such that $\sup_w \mathbb{E}_F[u_P(w; \epsilon)] = x^* - c(x^*) + c'(x^*) \cdot a$.

Let F be the exponential distribution with rate $\lambda = -\frac{1}{a}$ on the range $[a, +\infty)$. We have $F(\epsilon) = 1 - e^{-\lambda(\epsilon - a)}$, and its density is $f(\epsilon) = \lambda e^{-\lambda(\epsilon - a)}$. We check that the mean of ϵ is equal to $\frac{1}{\lambda} + a$, which is equal to 0. We have $\frac{1-F(\epsilon)}{f(\epsilon)} = \frac{1}{\lambda} = -a$ for all ϵ, so the virtual value for type ϵ (the integrand in (3.2)) is equal to $x(\epsilon) - c(x(\epsilon)) + a \cdot c'(x(\epsilon))$. The maximizer of this virtual value is equal to x^*, which is independent of ϵ, so the optimal contract given F is a reliable contract that implements x^*. We obtain that

$$\sup_w \mathbb{E}_F[u_P(w; \epsilon)] = x^* - c(x^*) + c'(x^*) \cdot a.$$

We conclude that the maxmin profit is equal to $x^* - c(x^*) + c'(x^*) \cdot a$.

If a is close to 0, the payoff guarantee is close to the first-best surplus, which is positive. Moreover, decreasing a enlarges the set of possible distributions of ϵ; indeed, the distribution takes the range $[a, b]$, and decreasing a expands the range of ϵ. Hence the payoff from a reliable contract decreases as a decreases. Finally, as a goes to $-\infty$, we have x^* goes to 0 (otherwise $c'(x^*) \cdot a$ goes to $-\infty$), which means the maxmin profit $x^* - c(x^*) + c'(x^*) \cdot a$ goes to 0. □

To tie up loose ends, we should acknowledge a difference between the original formulation of our problem and the screening formulation in this section. The original formulation required $w(y)$ to be specified for all possible y. Given a menu formulation of a contract, when we rewrite it in the original language, we need to deal with the possibility that the agent might want to choose an output level y that was not in the original menu, i.e. not equal to $y(\epsilon)$ for any ϵ. That is, we need to make sure w is extended to an upper semi-continuous function on the whole domain $[a, b + \overline{x}]$ without incentivizing any off-equilibrium choices of y. However, for Theorem 2.2 this is not a problem, since we have already described how to specify w on the whole domain (in Lemma 2.1).

(This potential issue is resolved in the same way in the model of the next section.)

4 Finite upper bound

Theorem 2.2 assumes that the noise ϵ has no upper bound (i.e. $b = +\infty$). This assumption allows us to clearly identify the worst-case distribution (i.e. an exponential distribution) and prove that the optimal contract is reliable. However, it might seem unnatural to assume that the lower bound a is finite, but the upper bound b is infinite. As a robustness check, we investigate whether the optimal contract is still reliable if b is finite.

In fact, the optimal contract is no longer reliable. In particular the "no distortion at the top" property, familiar from other screening models, will hold in our setting: The agent will be induced to choose the first-best effort level, $x^1 = \arg\max_x(x - c(x))$, at the extreme $\epsilon = b$, but not for lower values of ϵ. However, the optimal effort schedule is "somewhat" reliable if b is large enough. In particular, as b goes to infinity, the optimal effort schedule converges (locally uniformly) to the constant effort x^* implemented by the optimal reliable contract from Theorem 2.2.

In this section we consider a fixed, finite b, and derive the optimal contract. We will then show that it converges to a reliable contract as b goes to infinity. The derivation is tedious, so we first provide an intuitive description, and interested readers can take a look at the formal derivation.

Informal derivation To derive the optimal contract, we construct an effort schedule $x(\epsilon)$ such that the principal's payoff $u_P(w;\epsilon)$ is linear in ϵ. The linearity of $u_P(w;\epsilon)$ ensures that for any distribution of ϵ the principal earns an expected profit equal to $u_P(w;0)$. Moreover, given any contract $\widehat{w}$, we can imagine an alternative contract w such that $u_P(w;\epsilon)$ is linear in ϵ, and contract w guarantees the same profit as $\widehat{w}$. Hence it suffices to consider contracts w for which $u_P(w;\epsilon)$ is linear.

Figure 2 illustrates this argument. Contract $\widehat{w}$ yields a non-linear profit function $u_P(\widehat{w};\epsilon)$. Its worst-case expected profit, over all distributions of ϵ with mean zero, is given by the intercept of the thick black line with the vertical axis; the worst-case distribution puts all weight on ϵ_1 and ϵ_2. If we find a new contract w whose profit function is exactly this thick black line, then its worst-case expected profit is the same as the original contract. This motivates us, when solving for the optimal contract, to restrict attention to contracts w for which $u_P(w;\epsilon)$ is linear.

We write $u_P(w;\epsilon)$ from Equation (3.1): $u_P(w;\epsilon) = x(\epsilon) + \epsilon - c(x(\epsilon)) - \int_a^\epsilon c'(x(s)) \cdot ds$.

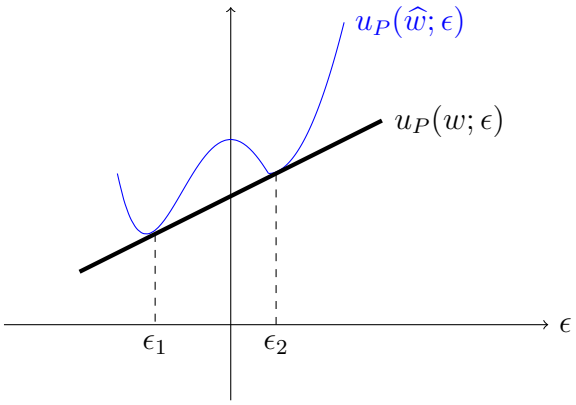

Figure 2: The profit guarantee of contract $\widehat{w}$ is equal to $u_P(w;0)$.

The first derivative is equal to

$$\frac{\partial}{\partial \epsilon} u_P(w;\epsilon) = x'(\epsilon) + 1 - c'(x(\epsilon)) \cdot x'(\epsilon) - c'(x(\epsilon)) = (x'(\epsilon)+1) \cdot (1 - c'(x(\epsilon))).$$

We construct the effort schedule from the following differential equation:

$$(x'(\epsilon)+1) \cdot (1 - c'(x(\epsilon))) = \beta, \tag{4.1}$$

for some constant $\beta > 0$, with the boundary condition that $x(b) = x^1$ is the first-best effort. This boundary condition is given by the "no distortion at the top property" in the integral (3.2).

The solution to (4.1) gives us the optimal effort schedule. We have that $x(\epsilon)$ is continuous, increasing, and convex. There is one subtle problem: as ϵ decreases, $x(\epsilon)$ might eventually become negative; in that case we choose the effort schedule $\max\{0, x(\epsilon)\}$. The formal proof shows that this is the correct optimal effort schedule. Figure 3 illustrates the two possible forms of $x(\epsilon)$. The left one comes directly from the differential equation (4.1), and the right one modifies the negative value of x to 0.

As b goes to infinity, convexity implies that the beginning part of $x(\epsilon)$ (for small ϵ) is close to flat. Therefore the optimal effort schedule is "approximately" reliable if b is large enough.[3]

[3]A rough intuition for convexity is as follows: suppose the contract induces higher effort at ϵ than at $\epsilon - \eta$, where η is small. Motivating this effort means that the pay per unit output must be higher at ϵ than at $\epsilon - \eta$. But the profit function should be linear, so to compensate for this extra pay, the incremental effort induced at $\epsilon + \eta$ should be higher still.

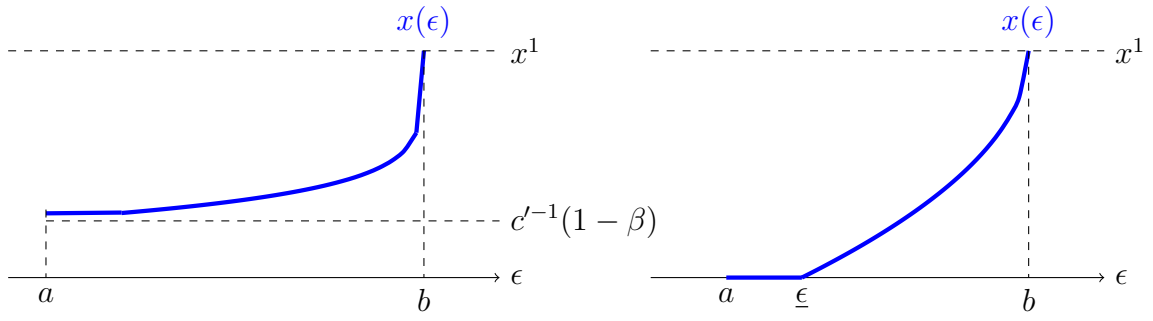

Figure 3: Construction of $x(\epsilon)$.

Formal derivation We first construct a family of candidate optimal contracts, parameterized by a real number $\alpha < 1$. (It is related to β from above by $\alpha = 1 - \beta$.) For each such α, we begin by constructing a function $x_\alpha(\epsilon)$, which describes the agent's effort level when the shock realization is ϵ, in the corresponding candidate contract.

To define x_α, we go through the following steps. First define function ϵ_α as follows. If $\alpha \geq 0$, the domain of ϵ_α is $((c')^{-1}(\alpha), x^1]$ (recall x^1 was defined as the first-best effort). If $\alpha < 0$, the domain of ϵ_α is $[0, x^1]$. On its domain we set

$$\epsilon_\alpha(x) = -\int_x^{x^1} \frac{1 - c'(s)}{c'(s) - \alpha}\, ds + b. \tag{4.2}$$

Note that $\epsilon_\alpha(x)$ is increasing and continuously differentiable. Define $x_\alpha(\epsilon)$ as the inverse of $\epsilon_\alpha(x)$. Then x_α is also increasing. Since $x_\alpha(\epsilon)$ is going to represent the agent's effort choice upon observing ϵ, we should make sure it is defined for all possible values of ϵ. For this purpose, we extend x_α downward to be defined on the domain $(-\infty, b]$, as follows. If $\alpha < 0$, then x_α as originally defined has domain $[\epsilon_\alpha(0), b]$, and we extend the domain by defining $x_\alpha(\epsilon) = 0$ for all $\epsilon < \epsilon_\alpha(0)$. Intuitively, domain extension adds a horizontal ray to the left of the graph of x_α, and this ray is at the minimal value of x_α. If $\alpha \geq 0$, then we have $\lim_{x \to (c')^{-1}(\alpha)} \epsilon_\alpha(x) = -\infty$.[4] We see that x_α already has the domain $(-\infty, b]$, so we don't need to extend it.

We know that x_α is non-decreasing and continuous on $(-\infty, b]$. To characterize the optimal contract, we also need to define $\underline{\epsilon}_\alpha$ as follows. If $\alpha < 0$ and $a < \epsilon_\alpha(0)$, then $\underline{\epsilon}_\alpha = \epsilon_\alpha(0)$; otherwise $\underline{\epsilon}_\alpha = a$. Notice that if $\underline{\epsilon}_\alpha \neq a$, then for all $\epsilon \in [a, \underline{\epsilon}_\alpha]$ we have

[4] Let $x_0 = (c')^{-1}(\alpha)$. For s sufficiently close to x_0 we have $c'(s) - \alpha < k(s - x_0)$, where k is any fixed positive number greater than $c''(x_0)$. Thus $\frac{1-c'(s)}{c'(s)-\alpha} = \frac{1-\alpha}{c'(s)-\alpha} - 1 > \frac{1-\alpha}{k(s-x_0)} - 1$ for s sufficiently close to x_0, say in the interval $(x_0, x_0 + \delta)$. The integral $\int_{x_0}^{x_0+\delta} \frac{1}{s-x_0}\, ds$ diverges, so $\lim_{x \to (c')^{-1}(\alpha)} \epsilon_\alpha(x) = -\infty$.

$x_\alpha(\epsilon) = 0$. The next proposition shows explicitly how to construct the optimal contract.

Proposition 4.1. *Suppose that $a > -\infty$ and $b < \infty$. Then:*

(a) For any real number $\alpha < 1$, define contract w_α such that

$$w_\alpha(x_\alpha(\epsilon) + \epsilon) = c(x_\alpha(\epsilon)) + \int_a^\epsilon c'(x_\alpha(s))\, ds \qquad \text{for every } \epsilon \in [a, b].$$

Moreover let $w_\alpha(y) \leq c(y - a)$ for all $y < x_\alpha(a) + a$, and $w_\alpha(y) = w_\alpha(x_\alpha(b) + b)$ for all $y > x_\alpha(b) + b$.

Then for some α the contract w_α maximizes V_P over all possible contracts, and the payoff from w_α is equal to $V_P(w_\alpha) = x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \alpha\underline{\epsilon}_\alpha$. For the value of α that maximizes this expression, the corresponding contract w_α is optimal.

(b) The optimal contract w_α is strictly better for the principal than any reliable contract. The optimal contract guarantees a positive payoff.

We defer the full proof of Proposition 4.1 to the appendix, but here is a summary. We cannot simply look for the worst-case F as in Theorem 2.2: for some choices of parameters, there will not exist any F for which the candidate optimal effort schedule x_α maximizes the virtual value pointwise. Thus, rather than look for a saddle point as before, we need a different approach.

We first consider any candidate optimal contract, and consider a linear bound as in the informal derivation above; let the slope of this line be $1 - \alpha$. Then we show that the optimal contract must induce effort bounded away from zero (Lemma C.2) in the interval $(\underline{\epsilon}_\alpha, b]$, so we can write the transfer rule as the integral envelope formula in Equation (3.1). We use this formula to show that the contract cannot give a better guarantee than w_α unless the surplus generated at the upper end $\epsilon = b$ is greater than under w_α — which is impossible, since w_α generates first-best surplus there.

We have two takeaways from this section. First, when b is finite, the effort schedule exhibits the "no distortion at the top" property in screening models. This is in line with other screening literature that has previously found that this property holds with bounded, but not with unbounded, type distributions; see e.g. the discussion of optimal income taxation in Saez (2001).

Second, the optimal effort schedule is convex and non-decreasing, so as b goes to infinity the initial part of the effort schedule (e.g. for small shocks) is nearly flat. That is, for any fixed finite $\widetilde{b}$, the effort schedule for $\epsilon \in [a, \widetilde{b}]$ must be nearly flat when b is

large. Moreover, its value on this interval must converge, as $b \to \infty$, to the effort induced by the optimal (reliable) contract for $b = \infty$ (that is, the x^* identified in the proof of Theorem 2.2). To see this convergence, notice that because profit is a linear function of ϵ, the guarantee from the optimal contract w equals $u_P(w; 0)$. From the envelope integral formula, this is close to the corresponding guarantee from a reliable contract that induces effort $x(0)$. So as $b \to \infty$, the effort level $x(0)$ in the optimal contract must converge to the effort level in the optimal reliable contract.

Note that because the principal's problem may not have a saddle point, the principal may be able to do strictly better by committing to a randomized contract, rather than using a deterministic contract as we have assumed (see Auster, 2016; Carrasco et al., 2015; Kos and Messner, 2015). We do not pursue this question further here.

5 Conclusion

We have considered a simple property of linear incentive contracts, which we termed *reliability*: they incentivize the same action by the agent, regardless of the situation (represented in our model by a privately observed shock to output). This property helps make linear contracts particularly easy to understand and analyze. Our concern, however, was with whether reliability can provide a microfoundation for linear contracts when the principal is ultimately concerned with profit. To model this, we aimed to write down a model in which such contracts are robust not only in terms of the action they induce, but also in terms of the expected profit, net of payment to the agent, that is guaranteed to the principal.

By adopting risk-neutrality and limited liability, we have set up the model with a simple tension — incentivizing more productive effort comes with an information-rent cost. And we have adopted a simple formulation of maxmin uncertainty about the shock distribution, to represent the desire for robustness. To cross the bridge from certainty about the action to certainty about expected net profit, we have further relied on two minimalist but important functional form assumptions: the additive specification for production, and the assumption on the principal's knowledge of the shock distribution (knowing the mean but nothing more). Together, these assumptions ensure that reliable contracts perfectly hedge the principal's uncertainty: Because the agent's effort is constant and the expected shock is known, expected output is certain; *and* because the contracts are linear in output over the relevant range, expected payment to the agent is also certain; thus, expected net profit is certain.

We see that reliable contracts are indeed optimal in such a model, but subtleties arise. First, even among reliable contracts, linear contracts are never optimal, since they give information rents at the bottom that can be avoided. Second, an asymmetry assumption is required: reliable contracts are optimal only when the possible shocks are unbounded above but bounded below. In particular, when shocks are bounded above, a situation that is familiar from screening applies: in general, incentivizing effort is costly in terms of information rent provided to higher types, so the principal wants to incentivize less-than-first-best effort; but for higher types this tradeoff disappears, and it is optimal to induce first-best effort at the top. We have shown how to explicitly describe the optimal contract in this situation.

How reasonable are our assumptions? In particular, for the uncertainty about the shock distribution (which is the unconventional part of the model), is it reasonable to assume that the principal knows bounds a, b, and the mean of the shock, but nothing more? One methodological defense is as follows: given that we wish to express robustness by incorporating uncertainty about the distribution, these are about the simplest minimal assumptions one can make to place some structure on the problem. Indeed, some other contemporary work on maxmin mechanism design (Brooks, 2013; Carrasco et al., 2015; Kos and Messner, 2015) makes similar assumptions of known mean and bounds. The asymmetry of finite a but infinite b might be justified if we imagine that there is a natural lower bound to output, but that the project undertaken by the agent could potentially be wildly successful.

The basic idea that it is optimal to induce constant effort (independent of the agent's private information) for reasons of simplicity, as in Edmans and Gabaix (2011), has intuitive appeal. One positive interpretation of our results is as a modeling handbook: if a theorist wishes to write down an agency model that reflects this intuitive appeal, we can provide quite specific guidance on how to do so. A negative interpretation would be that this reliability argument seems to depend on a very particular combination of assumptions. Perhaps there are still more general-purpose reasons to favor reliable contracts on account of their simplicity, but if so, a more involved model is needed to express them.

A Proof of Lemma 2.1

Proof of Lemma 2.1. We first prove the "if" part. Suppose w satisfies the conditions given in Lemma 2.1. We claim that the agent always chooses effort x^*. For all ϵ, the agent gets utility $w(x^* + \epsilon) - c(x^*)$ from exerting effort x^*. By definition of w, we have

$w(x^*+\epsilon)-c(x^*)=c'(x^*)(x^*+\epsilon-a)+h-c'(x^*)x^*=c'(x^*)(\epsilon-a)+h$. We show that no other effort yields more than $c'(x^*)(\epsilon-a)+h$. Suppose the agent observes ϵ and chooses effort x. Let $y=x+\epsilon$. We have three cases.

- If $y\in[a,a+x^*]$, then $w(x+\epsilon)-c(x)=w(y)-c(y-\epsilon)\le c(y-a)+h-c(y-\epsilon)$. Since c is increasing and convex, we have $c(y-a)+h-c(y-\epsilon)\le c'(y-a)(\epsilon-a)+h\le c'(x^*)(\epsilon-a)+h\le w(x^*+\epsilon)-c(x^*)$. Hence the agent exerts effort x^*.

- If $y\in[a+x^*,b+x^*]$, then $w(x+\epsilon)-c(x)=w(y)-c(y-\epsilon)=c'(x^*)(y-x^*-a)+h+c(x^*)-c(y-\epsilon)=[c'(x^*)y-c(y-\epsilon)]-c'(x^*)(x^*+a)+h+c(x^*)$. Since c is convex, the term $c'(x^*)y-c(y-\epsilon)$ is concave in y. The first order condition yields $c'(x^*)=c'(y-\epsilon)$, which means the term is maximized when $y=x^*+\epsilon$. Thus in this interval x^* is the optimal effort level.

- If $y\in[b+x^*,b+\overline{x}]$, then $w(x+\epsilon)-c(x)=w(y)-c(y-\epsilon)\le c(y-b)+c'(x^*)(b-a)+h-c(y-\epsilon)=[c(y-b)-c(y-\epsilon)]+c'(x^*)(b-a)+h$. The term $c(y-b)-c(y-\epsilon)$ is decreasing, so the maximum takes place when $y=b+x^*$, which brings us back to the previous case.

We now prove the "only if" part. Suppose a contract always induces x^*. Since $x^*>0$, the envelope integral formula holds (see the beginning of Section 3), so for all ϵ we have

$$w(x^*+\epsilon)=c(x^*)+U(a)+\int_a^\epsilon c'(x^*)\,ds=c(x^*)+U(a)+(\epsilon-a)c'(x^*).$$

Let $h=U(a)$. We have $h\ge 0$ because w satisfies the limited liability constraint. Thus, contract w satisfies $w(y)=c'(x^*)(y-a-x^*)+h+c(x^*)$ for all $y\in[x^*+a,x^*+b]$.

If $y\in[a,a+x^*]$, let $\epsilon=a$, we get $w(y)-c(y-a)\le w(x^*+a)-c(x^*)=h$, so $w(y)\le c(y-a)+h$. If $y\in[b+x^*,b+\overline{x}]$, let $\epsilon=b$, we get $w(y)-c(y-b)\le w(x^*+b)-c(x^*)=c'(x^*)(b-a)+h$, so $w(y)\le c(y-b)+c'(x^*)(b-a)+h$.

Finally, it is always incentive-compatible for the agent to choose effort x^*. If he does so, expected output is x^* and the principal's expected payment is $\mathbb{E}[w(y)|y\in[a+x^*,b+x^*]]=c'(x^*)(x^*-a)+h+c(x^*)-c'(x^*)x^*$. Thus, for any distribution of ϵ, the principal's expected payoff is $x^*+c'(x^*)a-c(x^*)-h$.

Note that there may sometimes exist other effort levels the agent is willing to choose, if $\epsilon=a$ and the constraint $w(y)\le c(y-a)+h$ holds with equality at several values of y (and likewise at $\epsilon=b$). Consequently, for some distributions of ϵ the principal may be able to get a higher expected payoff than indicated above. However, if $a<\epsilon<b$ then

the agent's *unique* optimal choice of effort is x^* (we can see this from the second bullet point, using the strict convexity of c). So by considering the distribution F that puts probability 1 on $\epsilon = 0$, we conclude that $V_P(w) = x^* + c'(x^*)a - c(x^*) - h$ exactly. □

B Properties of w_α and x_α

In this section we study the contract w_α, which has been defined in Section 4 for fixed, finite values of a and b. We first prove Lemma B.1, which pins down $x_\alpha(\epsilon)$. Then we prove a lemma (B.2) which computes the profit guarantee of w_α. The lemma also shows that some contract of the form w_α performs strictly better than any reliable contract.

We clarify the order of lemma dependency to avoid any impression that we are using circular reasoning. The proof of Lemma B.1 is self-contained, and the proof of Lemma B.2 cites Lemma B.1. Proofs of lemmas in later sections (C and D) will depend on lemmas in this section.

Lemma B.1. *For all $\epsilon \geq \underline{\epsilon}_\alpha$ we have*

$$x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) = [x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))] + \int_{\underline{\epsilon}_\alpha}^{\epsilon} (c'(x_\alpha(s)) - \alpha)\, ds.$$

Proof of Lemma B.1. We derive a differential equation for x_α. Whenever x is in the interior of the domain of ϵ_α, we have

$$\begin{aligned} \epsilon'_\alpha(x) &= \frac{1 - c'(x)}{c'(x) - \alpha} \\ 1 + \frac{1}{\epsilon'_\alpha(x)} &= \frac{1 - \alpha}{1 - c'(x)}. \end{aligned}$$

Therefore the inverse function $x_\alpha(\epsilon)$ is differentiable for $\epsilon > \underline{\epsilon}_\alpha$, and its derivative satisfies

$$\begin{aligned} 1 + x'_\alpha(\epsilon) &= \frac{1 - \alpha}{1 - c'(x_\alpha(\epsilon))} \\ (1 - c'(x_\alpha(\epsilon)))x'_\alpha(\epsilon) &= c'(x_\alpha(\epsilon)) - \alpha. \end{aligned}$$

Since x_α is continuous at $\underline{\epsilon}_\alpha$, we obtain that for all $\epsilon \geq \underline{\epsilon}_\alpha$:

$$x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) = [x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))] + \int_{\underline{\epsilon}_\alpha}^{\epsilon} (c'(x_\alpha(s)) - \alpha)\, ds.$$

□

Lemma B.2. *The contract w_α induces effort schedule x_α. The profit guarantee of w_α satisfies $V_P(w_\alpha) \geq x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \underline{\epsilon}_\alpha\alpha$, with equality unless $\underline{\epsilon}_\alpha \geq 0$, in which case $V_P(w_\alpha) = 0$. Moreover we have $\sup_\alpha V_P(w_\alpha) > \max_x (x - c(x) + c'(x)a)$.*

Proof. Let $y_\alpha(\epsilon) = x_\alpha(\epsilon) + \epsilon$. Since x_α is continuous and non-decreasing, y_α is continuous and increasing. If $y < y_\alpha(a)$, then $w_\alpha(y) - c(y - \epsilon) \leq w_\alpha(y_\alpha(a)) - c(x_\alpha(a)) + c(y - a) - c(y - \epsilon) < w_\alpha(y_\alpha(a)) - c(y_\alpha(a) - \epsilon)$, so the agent would rather produce $y_\alpha(a)$. On the other hand, if $y > y_\alpha(b)$, then $w_\alpha(y) - c(y - \epsilon) < w_\alpha(y_\alpha(b)) - c(y_\alpha(b) - \epsilon)$, so the agent would rather produce $y_\alpha(b)$. Thus we can assume that the agent produces an output in the interval $[y_\alpha(a), y_\alpha(b)]$. Since w_α satisfies the envelope integral formula, and y_α is increasing, w_α implements y_α.

We can calculate the payoff from w_α using Lemma B.1. For all $\epsilon \geq \underline{\epsilon}_\alpha$, we have

$$\begin{aligned} x_\alpha(\epsilon) - w_\alpha(x_\alpha(\epsilon) + \epsilon) &= x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) - \int_a^\epsilon c'(x_\alpha(s))\, ds \\ &= x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) - \int_{\underline{\epsilon}_\alpha}^\epsilon c'(x_\alpha(s))\, ds \\ &= x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) - \int_{\underline{\epsilon}_\alpha}^\epsilon [c'(x_\alpha(s)) - \alpha]\, ds - (\epsilon - \underline{\epsilon}_\alpha)\alpha \\ &= x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) - (\epsilon - \underline{\epsilon}_\alpha)\alpha. \end{aligned}$$

If $\underline{\epsilon}_\alpha = a$, then it follows that $V_P(w_\alpha) = x_\alpha(a) - c(x_\alpha(a)) + a\alpha$.

If $\underline{\epsilon}_\alpha \neq a$ (which can only happen when $\alpha < 0$), then $x_\alpha(\underline{\epsilon}_\alpha) = 0$ and the right-hand side above simplifies: it becomes $-(\epsilon - \underline{\epsilon}_\alpha)\alpha$ for any realization of ϵ that is $\geq \underline{\epsilon}_\alpha$. Note that this quantity is ≥ 0 since $\alpha < 0$. On the other hand, whenever $\epsilon < \underline{\epsilon}_\alpha$, $x(\epsilon) = w(y(\epsilon)) = 0$, so the expression above is just 0. Hence, the principal's payoff for any ϵ is

$$\epsilon + x_\alpha(\epsilon) - w(x_\alpha(\epsilon) + \epsilon) = \epsilon + \max\{-(\epsilon - \underline{\epsilon}_\alpha)\alpha, 0\}.$$

This is a convex function, so the worst possible expected payoff over all mean-zero distributions F is just its value at $\epsilon = 0$: $\max\{\underline{\epsilon}_\alpha\alpha, 0\}$. So we indeed have the inequality $V_P(w_\alpha) \geq \underline{\epsilon}_\alpha\alpha = x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \underline{\epsilon}_\alpha\alpha$, with equality if $\underline{\epsilon}_\alpha < 0$.

Let x^* denote the effort that maximizes the payoff from a reliable contract. If $x^* > 0$ then the first-order condition is $1 - c'(x^*) + c''(x^*)a = 0$. Since $c''(x^*) > 0$, we deduce that $x^* < x^1$. And if $x^* = 0$ then of course we also have $x^* < x^1$. Let $\alpha = c'(x^*)$. Since $\alpha \geq 0$, we have $x_\alpha(a) > (c')^{-1}(\alpha) = x^*$. We also know that $x_\alpha(a) < x^1$, so we get $V_P(w_\alpha) = x_\alpha(a) - c(x_\alpha(a)) + a\alpha > x^* - c(x^*) + a\alpha = \max_x (x - c(x) + c'(x)a)$. □

C Lemmas needed for the Envelope Integral Formula

In this section we prove two requisite lemmas to ensure that the envelope integral formula holds. Recall from Section 3 that the transfer rule may not be unique if $x(\epsilon) = 0$ for some ϵ. We overcome this problem by showing that if a contract performs better than the contracts we identified in Theorem 2.2 and Proposition 4.1, then $x(\epsilon)$ is bounded strictly above 0 in our domain of interest.

The proof of Lemma C.1 is self-contained, and the proof of Lemma C.2 cites Lemma B.2.

Lemma C.1. *If $b = \infty$ and $V_P(w) > \max_x (x - c(x) + c'(x)a)$, then $\inf_\epsilon x(\epsilon) > 0$.*

Proof. We know that, for any distribution F over ϵ with mean 0, the expected value of $u_P(w;\epsilon)$ is at least $V_P(w)$ (this is the definition of $V_P(w)$). So if we consider the set $S \subseteq \mathbb{R}^2$ consisting of the convex hull of points $\{(\epsilon, u_P(w;\epsilon))|\epsilon \in [a,b]\}$, and the set $T = \{(0,\zeta)|\zeta < V_P(w)\}$, these two sets are disjoint. By the Separating Hyperplane Theorem, there exists a line separating them; taking $1-\alpha$ to be the slope of this line, we get $y(\epsilon) - w(y(\epsilon)) \geq (1-\alpha)\epsilon + V_P(w)$ for all $\epsilon \in [a,b]$.

If $\alpha < 0$, then the right hand side of this inequality grows at a rate $(1-\alpha)\epsilon$ as $\epsilon \to \infty$, but the left hand side is bounded above by $\bar{x} + \epsilon$. Hence we must have $\alpha \geq 0$.

Next, if $\alpha \geq 1$, then $y(a) - w(y(a)) \geq V_P(w)$. On the other hand, we always have $w(y(a)) - c(x(a)) \geq 0$ (since the agent can guarantee himself zero payoff at $\epsilon = a$ by taking zero effort); this implies $y(a) - w(y(a)) \leq a + x(a) - c(x(a)) \leq a + x^1 - c(x^1)$. We deduce that $V_P(w) \leq x^1 - c(x^1) + a$, which is the profit from a reliable contract that induces x^1, contradicting the assumption that $V_P(w) > \max_x (x - c(x) + c'(x)a)$.

We now restrict attention to $\alpha \in [0,1)$. Suppose $V_P(w) = \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha)) + \alpha a$, where δ is a positive real number. We have $y(\epsilon) - w(y(\epsilon)) \geq \delta + (1-\alpha)\epsilon + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha)) + \alpha a$. Consequently we get

$$x(\epsilon) - w(y(\epsilon)) \geq \delta + \alpha(a - \epsilon) + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha)). \tag{C.1}$$

For $\epsilon = a$, we again use $w(y(a)) - c(x(a)) \geq 0$, to infer that $x(a) - c(x(a)) \geq \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha))$. Let $\widehat{x}$ denote the effort level below x^1 for which $\widehat{x} - c(\widehat{x}) = \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha))$. We have $x(a) \geq \widehat{x} > 0$.

We claim that for all $\epsilon > a$, the following inequalities hold:

$$\begin{aligned} x(\epsilon) - c(x(\epsilon)) &\geq \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha)) \\ w(y(\epsilon)) - c(x(\epsilon)) &\geq \alpha(\epsilon - a). \end{aligned}$$

Let $\Delta = \widehat{x} - (c')^{-1}(\alpha) > 0$. We first prove these inequalities for $\epsilon \in [a, a+\Delta]$. Any such type ϵ is able to produce output $y(a)$, since $y(a) = a + x(a) \geq a + \widehat{x} \geq a + \Delta$. Consequently, the incentive constraint implies that $w(y(\epsilon)) - c(x(\epsilon)) \geq w(y(a)) - c(y(a) - \epsilon) \geq c(x(a)) - c(x(a) - (\epsilon - a))$. Since $x(a) - (\epsilon - a) \geq \widehat{x} - \Delta = (c')^{-1}(\alpha)$, the convexity of c implies that $c(x(a)) - c(x(a) - (\epsilon - a)) \geq \alpha(\epsilon - a)$. We thus obtained that $w(y(\epsilon)) - c(x(\epsilon)) \geq \alpha(\epsilon - a)$. Consequently, we have $x(\epsilon) - c(x(\epsilon)) = [x(\epsilon) - w(y(\epsilon))] + [w(y(\epsilon)) - c(x(\epsilon))] \geq \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha))$. Therefore the two inequalities hold for all $\epsilon \in [a, a + \Delta]$.

Suppose the two inequalities hold for the interval $[a + (k-1)\Delta, a + k\Delta]$. Then, the first inequality applied to $\epsilon = a + k\Delta$ ensures $x(a + k\Delta) \geq \widehat{x} > \Delta$, so $y(a + k\Delta) \geq a + (k+1)\Delta$, and hence any type $\epsilon \in [a + k\Delta, a + (k+1)\Delta]$ is able to produce output $y(a + k\Delta)$. Moreover, $y(a + k\Delta) - \epsilon \geq \widehat{x} + (a + k\Delta) - \epsilon \geq \widehat{x} - \Delta = (c')^{-1}(\alpha)$.

So, for any ϵ in this new interval, the incentive constraints yield

$$\begin{aligned} w(y(\epsilon)) - c(x(\epsilon)) &\geq w(y(a + k\Delta)) - c(y(a + k\Delta) - \epsilon) \\ &\geq \alpha k\Delta + c(x(a + k\Delta)) - c(y(a + k\Delta) - \epsilon) \quad \text{(the inductive hypothesis)} \\ &\geq \alpha k\Delta + c'(y(a + k\Delta) - \epsilon) \cdot (\epsilon - a - k\Delta) \quad \text{(the convexity of } c\text{)} \\ &\geq \alpha k\Delta + \alpha(\epsilon - a - k\Delta) \\ &= \alpha(\epsilon - a). \end{aligned}$$

As a result we have $x(\epsilon) - c(x(\epsilon)) \geq \delta + (c')^{-1}(\alpha) - c((c')^{-1}(\alpha))$.

Now, by induction, our two inequalities hold for all ϵ. Therefore, for all ϵ we have $x(\epsilon) \geq \widehat{x} > 0$. □

Lemma C.2. *Suppose $b < \infty$ and $V_P(w) > \sup_\alpha V_P(w_\alpha)$. Then there exists an $\alpha < 1$ such that $y(\epsilon) - w(y(\epsilon)) \geq (1 - \alpha)\epsilon + V_P(w)$ for all ϵ, and $\inf_{\epsilon \geq \underline{\epsilon}_\alpha} x(\epsilon) > 0$.*

Proof. By an application of the Separating Hyperplane Theorem, identical to that used in Lemma C.1, we know that there is an α such that $y(\epsilon) - w(y(\epsilon)) \geq (1 - \alpha)\epsilon + V_P(w)$ for all ϵ.

If $\alpha \geq 1$, then $V_P(w) \leq y(a) - w(y(a)) = a + x(a) - w(y(a)) \leq a + x(a) - c(x(a)) \leq x^1 - c(x^1) + a$. By Lemma B.2 we have $x^1 - c(x^1) + a < \sup_\alpha V_P(w_\alpha)$, contradicting the

assumption that $V_P(w) > \sup_\alpha V_P(w_\alpha)$.

If $\alpha < 1$, then for some $\delta > 0$ we have $y(\epsilon) - w(y(\epsilon)) \geq \delta + (1-\alpha)\epsilon + V_P(w_\alpha) \geq \delta + (1-\alpha)\epsilon + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \alpha\underline{\epsilon}_\alpha$ for all ϵ. We get

$$x(\epsilon) - w(y(\epsilon)) \geq \delta + \alpha(\underline{\epsilon}_\alpha - \epsilon) + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)).$$

If $\alpha \leq 0$, then for all $\epsilon \geq \underline{\epsilon}_\alpha$ the right hand side is at least $\delta + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))$, and the left hand side is at most $x(\epsilon) - c(x(\epsilon))$, which implies that $x(\epsilon) \geq \delta > 0$ for all $\epsilon \geq \underline{\epsilon}_\alpha$. On the other hand, if $\alpha > 0$, then $x_\alpha(\underline{\epsilon}_\alpha) \geq (c')^{-1}(\alpha)$, and we can proceed exactly the same way as Lemma C.1 (starting from expression (C.1) and replacing a with $\underline{\epsilon}_\alpha$). We again obtain that $x(\epsilon) \geq \widehat{x} > 0$ for all $\epsilon \geq \underline{\epsilon}_\alpha$. □

Lemma C.2 allows us to use the envelope integral formula when $\epsilon \geq \underline{\epsilon}_\alpha$:

$$w(y(\epsilon)) = c(x(\epsilon)) + U(\underline{\epsilon}_\alpha) + \int_{\underline{\epsilon}_\alpha}^{\epsilon} c'(x(s))\, ds.$$

We will use this formula in the proof of Proposition 4.1.

D Proof of Proposition 4.1

Proof of Proposition 4.1. First we know from Lemma B.2 that contract w_α induces effort schedule x_α, and that the payoff satisfies $V_P(w_\alpha) \geq x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \alpha\underline{\epsilon}_\alpha$, with equality if $\underline{\epsilon}_\alpha < 0$.

We now show that the principal cannot achieve a payoff guarantee of more than $\sup_\alpha V_P(w_\alpha)$. Suppose for contradiction that there exists a contract w such that $V_P(w) > \sup_\alpha V_P(w_\alpha)$. Lemma C.2 implies that there exists an $\alpha < 1$ such that $y(\epsilon) - w(y(\epsilon)) \geq (1-\alpha)\epsilon + V_P(w)$ for all $\epsilon \in [a, b]$. Moreover, $\inf_{\epsilon \geq \underline{\epsilon}_\alpha} x(\epsilon) > 0$.

Thus there exists a $\delta > 0$ such that $x(\epsilon) + \epsilon - w(y(\epsilon)) \geq \delta + (1-\alpha)\epsilon + V_P(w_\alpha)$ for all $\epsilon \in [a, b]$. For all $\epsilon \geq \underline{\epsilon}_\alpha$, the envelope integral formula implies that

$$w(y(\epsilon)) = c(x(\epsilon)) + U(\underline{\epsilon}_\alpha) + \int_{\underline{\epsilon}_\alpha}^{\epsilon} c'(x(s))\, ds.$$

We deduce that

$$\begin{aligned} x(\epsilon) - c(x(\epsilon)) &\geq \delta - \alpha\epsilon + \int_{\underline{\epsilon}_\alpha}^{\epsilon} c'(x(s))\, ds + V_P(w_\alpha) \\ &\geq \delta - \alpha\epsilon + \int_{\underline{\epsilon}_\alpha}^{\epsilon} c'(x(s))\, ds + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha)) + \alpha\underline{\epsilon}_\alpha \\ &= [\delta + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))] + \int_{\underline{\epsilon}_\alpha}^{\epsilon} (c'(x(s)) - \alpha)\, ds. \end{aligned}$$

On the other hand, for all $\epsilon \geq \underline{\epsilon}_\alpha$ we have

$$x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) = [x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))] + \int_{\underline{\epsilon}_\alpha}^{\epsilon} (c'(x_\alpha(s)) - \alpha)\, ds$$

by Lemma B.1. We obtain that

$$x(\epsilon) - c(x(\epsilon)) \geq [\delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon))] + \int_{\underline{\epsilon}_\alpha}^{\epsilon} (c'(x(s)) - c'(x_\alpha(s)))\, ds.$$

We claim that $x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon))$ for all $\epsilon \in [\underline{\epsilon}_\alpha, b]$. This inequality would give us a contradiction because $x_\alpha(b)$ is the first-best effort.

Taking $\epsilon = \underline{\epsilon}_\alpha$, we get $x(\underline{\epsilon}_\alpha) - c(x(\underline{\epsilon}_\alpha)) \geq \delta + x_\alpha(\underline{\epsilon}_\alpha) - c(x_\alpha(\underline{\epsilon}_\alpha))$. If $x(\epsilon) \geq x_\alpha(\epsilon)$ for all ϵ, then $c'(x(s)) - c'(x_\alpha(s)) \geq 0$ for all s, and we obviously have $x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon))$ because the integral term is always positive. Suppose on the contrary that there is an ϵ for which $x(\epsilon) < x_\alpha(\epsilon)$. Let $\epsilon^* = \inf\{\epsilon | x(\epsilon) < x_\alpha(\epsilon)\}$. Taking $\epsilon \to \epsilon^{*+}$ we have

$$x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) + \int_{\underline{\epsilon}_\alpha}^{\epsilon^*} (c'(x(s)) - c'(x_\alpha(s)))\, ds + \int_{\epsilon^*}^{\epsilon} (c'(x(s)) - c'(x_\alpha(s)))\, ds.$$

If $\epsilon^* = \underline{\epsilon}_\alpha$, then $x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon)) + \int_{\epsilon^*}^{\epsilon} (c'(x(s)) - c'(x_\alpha(s)))ds$. However, as $\epsilon \to \epsilon^*$, the integral converges to 0, so we have $x(\epsilon) - c(x(\epsilon)) > x_\alpha(\epsilon) - c(x_\alpha(\epsilon))$. Since $x_\alpha(\epsilon) \leq x^1$, we get $x(\epsilon) > x_\alpha(\epsilon)$ for ϵ close enough to ϵ^*, contradicting the definition of ϵ^*. Now suppose $\epsilon^* > \underline{\epsilon}_\alpha$. Since $x(s) \geq x_\alpha(s)$ for all $s < \epsilon^*$, we know that $x(s) - c(x(s)) \geq \delta + x_\alpha(s) - c(x_\alpha(s))$ for all $s < \epsilon^*$, which implies that $x(s) > x_\alpha(s)$ for all $s < \epsilon^*$. We deduce that the first integral $\int_{\underline{\epsilon}_\alpha}^{\epsilon^*} (c'(x(s)) - c'(x_\alpha(s)))\, ds$ is positive (note that we cannot have $\epsilon^* = \underline{\epsilon}_\alpha$). We also note that the second integral $\int_{\epsilon^*}^{\epsilon} (c'(x(s)) - c'(x_\alpha(s)))\, ds$ goes to 0 as $\epsilon \to \epsilon^{*+}$. Hence as $\epsilon \to \epsilon^{*+}$, we have $x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon))$, and again we get $x(\epsilon) > x_\alpha(\epsilon)$ for ϵ close enough to ϵ^*, contradicting the definition of

ϵ^*. Therefore we must have $x(\epsilon) - c(x(\epsilon)) \geq \delta + x_\alpha(\epsilon) - c(x_\alpha(\epsilon))$ for all $\epsilon \in [\underline{\epsilon}_\alpha, b]$. As mentioned above, we now reach a contradiction because $x_\alpha(b)$ is the first-best effort.

The above arguments show that no contract can guarantee a profit greater than $\sup_\alpha V_P(w_\alpha)$. Since $\underline{\epsilon}_\alpha$ and $x_\alpha(\underline{\epsilon}_\alpha)$ are continuous in α, the sup is achieved by some $\alpha \leq 1$. Furthermore, Lemma B.2 tells us that $\sup_\alpha V_P(w_\alpha) > \max_x (x - c(x) + c'(x)a)$. Hence a reliable contract performs strictly worse than some contract of the form w_α.

This tells us that the sup over α is attained in the interior — it cannot be attained at $\alpha = 1$ because this would correspond to a reliable contract (implementing first-best effort) which we know is not optimal. And it also tells us that the optimal contract yields a strictly positive payoff, since the zero-contract is reliable and gives profit 0. This completes the proof of the theorem. □

References

S. Auster. Robust contracting under common value uncertainty. Unpublished paper, 2016.

S. Bhattacharyya and F. Lafontaine. Double-sided moral hazard and the nature of share contracts. *The RAND Journal of Economics*, 26(4):761–781, 1995.

B. Biais, T. Mariotti, G. Plantin, and J.-C. Rochet. Dynamic security design: Convergence to continuous time and asset pricing implications. *The Review of Economic Studies*, 74 (2):345–390, 2007.

B. A. Brooks. Surveying and selling: Belief and surplus extraction in auctions. Unpublished paper, 2013.

V. Carrasco, V. F. Luz, P. Monteiro, and H. Moreira. Robust selling mechanisms. Unpublished paper, 2015.

G. Carroll. Robustness and linear contracts. *The American Economic Review*, 105(2): 536–563, 2015.

S. Chassang. Calibrated incentive contracts. *Econometrica*, 81(5):1935–1971, 2013.

L. Y. Chu and D. E. M. Sappington. Simple cost-sharing contracts. *The American Economic Review*, 97(1):419–428, 2007.

K.-S. Chung and J. C. Ely. Foundations of dominant strategy mechanisms. *Review of Economic Studies*, 74(2):447–476, 2007.

P. M. DeMarzo and M. J. Fishman. Optimal long-term financial contracting. *Review of Financial Studies*, 20(6):2079–2128, 2007.

A. Edmans and X. Gabaix. Tractability in incentive contracting. *Review of Financial Studies*, 24(9):2865–2894, 2011.

A. Edmans, X. Gabaix, T. Sadzik, and Y. Sannikov. Dynamic CEO compensation. *Journal of Finance*, 67(5):1603–1647, 2012.

A. Frankel. Aligned delegation. *The American Economic Review*, 104(1):66–83, 2014.

D. Garrett and A. Pavan. Managerial turnover in a changing world. *Journal of Political Economy*, 120(5):879–925, 2012.

D. Garrett and A. Pavan. Dynamic managerial compensation: A variational approach. *Journal of Economic Theory*, 159:775–818, 2015.

D. F. Garrett. Robustness of simple menus of contracts in cost-based procurement. *Games and Economic Behavior*, 87:631–641, 2014.

B. Holmström and P. Milgrom. Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, 55(2):303–328, 1987.

L. Hurwicz and L. Shapiro. Incentive structures maximizing residual gain under incomplete information. *The Bell Journal of Economics*, 9(1):180–191, 1978.

N. Kos and M. Messner. Selling to the mean. Unpublished paper, 2015.

J. M. Lacker and J. A. Weinberg. Optimal contracts under costly state falsification. *The Journal of Political Economy*, 97(6):1345–1363, 1989.

J.-J. Laffont and J. Tirole. Using cost observation to regulate firms. *Journal of Political Economy*, 94(3):614–641, 1986.

G. Lewis and P. Bajari. Moral hazard, incentive contracts, and risk: evidence from procurement. *The Review of Economic Studies*, 81(3):1201–1228, 2014.

J. M. López-Cuñat. Adverse selection under ignorance. *Economic Theory*, 16(2):379–399, 2000.

R. P. McAfee and J. McMillan. Competition for agency contracts. *The RAND Journal of Economics*, 18(2):296–307, 1987.

W. P. Rogerson. Simple menus of contracts in cost-based procurement and regulation. *The American Economic Review*, 93(3):919–926, 2003.

E. Saez. Using elasticities to derive optimal income tax rates. *The Review of Economic Studies*, 68(1):205–229, 2001.