Broad Validity of the First-Order Approach in Moral Hazard*

Eduardo M. Azevedo[†] Ilan Wolff[‡]

First version: March 7, 2025
This version: March 7, 2025

Abstract

The first-order approach (FOA) is the main tool in the study of the pure moral hazard principal-agent problem. Although many existing results rely on the FOA, its validity has been established only under relatively restrictive assumptions. We contribute three main findings.

First, we demonstrate in a broad array of examples that the FOA frequently fails when the agent's reservation utility is low (such as in principal-optimal contracts). However, the FOA holds when the agent's reservation utility is at least moderately high (such as in competitive settings where agents receive high rents).

Second, our main theorem formally shows that the FOA is valid in a standard limited liability model when the agent's reservation utility is sufficiently high. The theorem also establishes existence and uniqueness of the optimal contract.

Third, we use the FOA to derive tractable optimal contracts across a broad array of settings. These contracts are both simple and intuitive, and under log utility, they are piecewise linear for numerous common output distributions (including Gaussian, exponential, binomial, Gamma, and Laplace).

*We are especially grateful to Lucas Maestri for originating key ideas in this paper. We also thank Humberto Moreira and Juuso Toikka for their valuable comments.

[†]Wharton: 3733 Spruce Street, Philadelphia, PA 19104. eazevedo@wharton.upenn.edu, www.eduardomazevedo.com.

[‡]Wharton: 3733 Spruce Street, Philadelphia, PA 19104. ijwolff@wharton.upenn.edu.

1 Introduction

One of the workhorse models in economics and related fields is the principal-agent problem with moral hazard. The principal hires the agent to take an action a in $\mathbb{R}_+$ that affects the distribution of output y. The principal can only condition payments on realized output, and designs a contract $w(y)$ to provide incentives to the agent. The agent chooses the action a to maximize her utility from wages minus her cost of effort,

$$\mathbb{E}\left[u(w(y))|a\right] - c(a),$$

and the principal chooses the contract to maximize expected profits.

The main solution method for this problem is the first-order approach, which assumes that only local deviations in a are binding. This yields a tractable, calculus-based formula for the optimal contract (see Holmström (1978) and the excellent survey Georgiadis (2022)). Many results in the literature rely on the first-order approach, even though Mirrlees (1999) (circulated in 1975) showed it is not always valid. In applied work, restrictive assumptions such as linear contracts or binary effort are often made to avoid this issue.[1]

Unfortunately, existing sufficient conditions for the first-order approach are restrictive. The seminal papers are Rogerson (1985) and Jewitt (1988), followed by an extensive literature.[2] Kadan, Reny, and Swinkels (2017) summarize the general view that "conditions facilitating the first-order approach are typically quite demanding". The key issues are elegantly explained by Chaigneau et al. (2022) and Conlon (2009). They explain that the sufficient conditions require contracts to be close to concave, precluding common contracts such as compensation with stock options.[3]

[1]Examples of papers assuming the first-order approach include, to cite a few, Jewitt et al. (2008); Moroni and Swinkels (2014); Chaigneau et al. (2022). Virtually all early work, including the seminal papers by Holmström (1978, 1979), and Zeckhauser (1970) assumes it. Holmström (1979) notes that "one has to assume that the agent's optimal choice of action is unique for the optimal [...] This assumption seems very difficult to validate [...] and regrettably we have to leave the question about its validity open." Edmans et al. (2009) is an example of an important recent paper using both binary effort and linear contracts. See the surveys Bolton and Dewatripont (2005); Salanié (2005); Georgiadis (2022) for further references.

[2]Important generalizations include Sinclair-Desgagné (1994); Conlon (2009); Jung and Kim (2015); Chade and Swinkels (2020), and Jung et al. (2024). Chaigneau et al. (2022, 2024) develop sufficient conditions with limited liability.

[3]Conlon (2009) explains the intuition clearly: "Unfortunately, the Jewitt conditions are tied to the concavity, not only of the technology, but also of the payment schedule. Thus, there are interesting cases where the Jewitt conditions should fail, since the payment schedule is often not concave. For example, managers often receive stock options, face liquidity constraints, [...] I believe that it is natural to expect the first-order approach itself to fail in such cases, since the agent's overall objective function will tend to be nonconcave [...]. The fact that CDFC and CISP imply concavity of the agent's payoff, regardless of the curvature of

Our main result, Theorem 1, shows that the first-order approach is broadly valid, as long as the agent's reservation utility is sufficiently high. The result holds even when contracts are option-like and the agent's problem has multiple local maxima. We consider a setting with limited liability, because of its practical importance and because it is known that limited liability or a similar assumption is needed to guarantee existence (Moroni and Swinkels, 2014). We find that, for low reservation utility, the standard view in the literature holds and the first-order approach often fails. However, for high reservation utility, we show that the first-order approach is satisfied. Thus, the issues with the validity of the first-order approach are likely to be important in examples with low reservation utility like when a monopsonist principal can impose highly unfavorable contracts on agents. When reservation utility is low, the first order approach is often invalid because it produces contracts where the agent has a global deviation to exert minimal effort. However, the first order approach is likely to be valid in settings with high reservation utility, such as competitive labor markets with high productivity workers. As reservation utility goes up, the first order approach's contract increases the incentive to work, eliminating global deviations, and restoring the first-order approaches validity.

Theorem 1 also shows that the optimal contract exists, is unique, and characterized by a simple formula. Section 6 show that optimal contracts are piecewise linear option contracts for many standard distributions and log agent utility, and gives simple computational agorithms. Thus, the limited liability model seems well suited for applied theory and empirical models that seek to predict the shape of optimal contracts $w(y)$.

Before going into the analysis, section 2 explains the gist of our results through simple examples, and reconciles our findings with the literature. Section 3 gives definitions and Section 4 states Theorem 1. Section 5 outlines the proof. Section 6 considers applications, extensions, and counter examples.

2 Examples and Literature Review

2.1 Examples Illustrating the Main Result

Consider the following **Gaussian-log utility example**. A risk-neutral principal hires an agent to work. The agent chooses action a at a cost $c(a)$ proportional to a^2. Output y,

[the payment], then suggests that the CDFC/CISP conditions are very restrictive, not that the first-order approach is widely applicable.".

which accrues to the principal, is normally distributed with mean a and standard deviation σ. The agent has some independent assets and log utility, so that her utility from a payment of $x \geq 0$ is $u(x) = \log(w_0 + x)$. For concreteness, suppose the principal is the owner of a small company and the agent is a professional manager. Illustrative values are a choices in the ballpark of \$100,000, standard deviation σ of \$20,000, and w_0 equal to \$50,000. The agent has reservation utility $\bar{U}$.

The principal designs a compensation contract with wage $w(y)$ to maximize profits. The optimal contract balances the goals of inducing effort, reducing the agent's risk, and providing the agent with her reservation utility. It is known that, without limited liability, an optimal contract does not exist. Mirrlees (1999) showed that it is possible for the principal to get arbitrarily close to her perfect-information payoff using contracts that impose harsh punishments with small probability. Thus, we consider the case of limited liability, requiring $w(y) \geq 0$.[4]

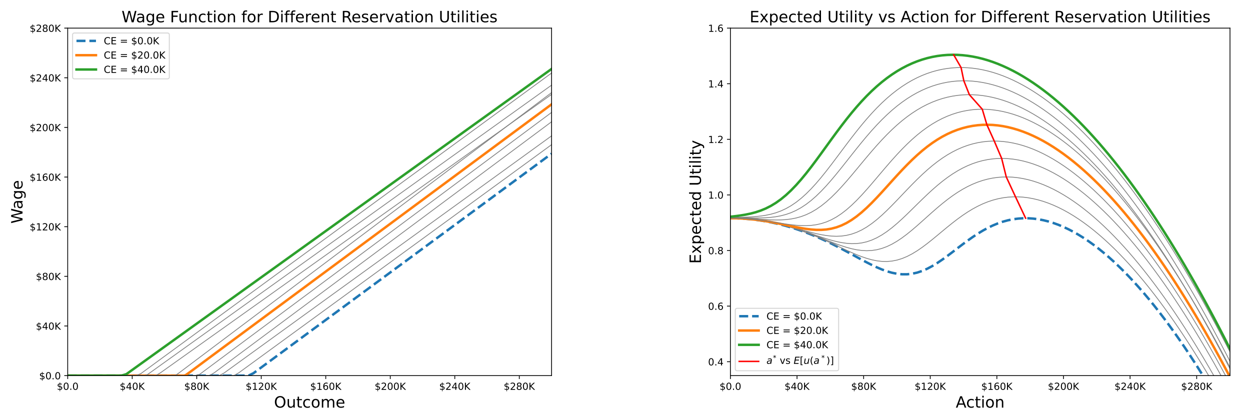
Figure 1 plots optimal contracts for a range of reservation utilities $\bar{U}$, with certain equivalents from \$0 to \$40,000. The left panels plot the optimal wage schedules $w(y)$. The optimal contracts are intuitive. They have a flat region to the left where, due to limited liability, the agent is paid $w(y) = 0$. And an increasing region to the right, which gives the agents incentives to work. As reservation utility increases, contracts become more generous. Wages schedules shift up, and the minimum output required to receive positive pay moves to the left.

The right panel is crucial for understanding when the first-order approach is valid. It plots the agent's expected utility $U(v^*, a)$ as a function of a, given the optimal contract. The first-order approach fails whenever there is a binding non-local deviation. In these cases, solving the problem with only the local incentive constraint leads to incorrect solutions, as pointed out by Mirrlees (1999). Cases in which the first-order approach fails are denoted with dashed lines.

The first-order approach fails spectacularly at the lowest reservation utility, with certain equivalent of \$0. The optimal contract is a tough bargain. The agent is expected to exert effort a equal to \$177.2K. The contract $w(y)$ specifies a payment of zero for any output lower than \$113.1K. Thus, the agent needs to produce relatively high output to get paid at all. This contract induces two local maxima in the function $U(v^*, a)$. The agent is indifferent between basically giving up with a local maximum at $a \approx 0$ and a local maxi-

[4]Moroni and Swinkels (2014) show that, without limited liability, existence often fails if the agent has negative infinite utility from a finite wage. Limited liability is a common and minimal way to avoid these types of existence issues.

Figure 1: Optimal Contracts with Gaussian Distribution and log Utility



(a) Canonical Contracts

(b) Expected Utility

Note: Panel a displays the wage function that solves the principal's profit maximization problem at different values of $\bar{U}$, and Panel b displays the agent's expected utility from a grid of possible actions given the corresponding wage functions from Panel a. The red line in Panel b plots the optimal action against opitmal utility for the agent. The values of $\bar{U}$ are expressed in terms of certainty equivalents. The agent's utility is $u(x) = log(\frac{x}{20K})$, the cost function is $c(a) = \theta a^2$, where $\theta = .01$, and the agent's starting wealth is $w_0 = 50K$. The output is $y \sim N(a, (20K^2))$, where a is the agent's action. Dotted lines indicate that the FOA is invalid. Our optimization routine has only one global IC constraint at $a = 1e - 3$. We conservatively say that the FOA is invalid if the Lagrange multiplier on the global IC constraint is greater than $1e - 5$ or if deviating to an action between 0 and .1 reduces the agent's utility by less than $1e - 2$.

mum at the recommended effort level $a = \$177.2K$. This is consistent with the standard view in the literature. Option-like contracts are natural, and thus it is difficult to find general sufficient conditions under which the first-order approach holds.

Surprisingly, the first-order approach becomes valid as soon as reservation utility is slightly higher. The constraint of giving the agent higher utility forces the principal to offer contracts with a higher value of $U(v^*, a)$ at the recommended action. This constraint moves the rightmost local maximum of $U(v^*, a)$ up. The rightmost maximum goes up faster than the leftmost local maximum. Thus, the rightmost maximum becomes the unique global maximum. Numerically, the first-order approach is valid for any reservation utility with certain equivalent above \$600, as seen in the figure.

Theorem 1 shows that this example is more general. The theorem shows that, under certain conditions, the first-order approach is always valid for sufficiently high reservation utility. The most substantial condition is that greater output is evidence of greater effort (monotone likelihood ratio). Figures 2 and 3 illustrate the generality of this point by plotting the same graphs in examples with different risk preferences and output distributions. All the examples display the same qualitative pattern, consistent with Theorem 1. We stress, however, that the monotone likelihood ratio assumption is crucial. For example, if output has a fat-tailed Student-t distribution, there are examples where the first-order approach fails for reservation utilities. The reason is that, positive and negative outliers are not very informative about effort, which generates global deviations.

Theorem 1 also guarantees existence and uniqueness of an optimal contract. Moreover, the solution has a simple calculus-based formula, as in the classic first-order approach literature following Holmström (1978). The Theorem shows that the optimal wage equals[5]

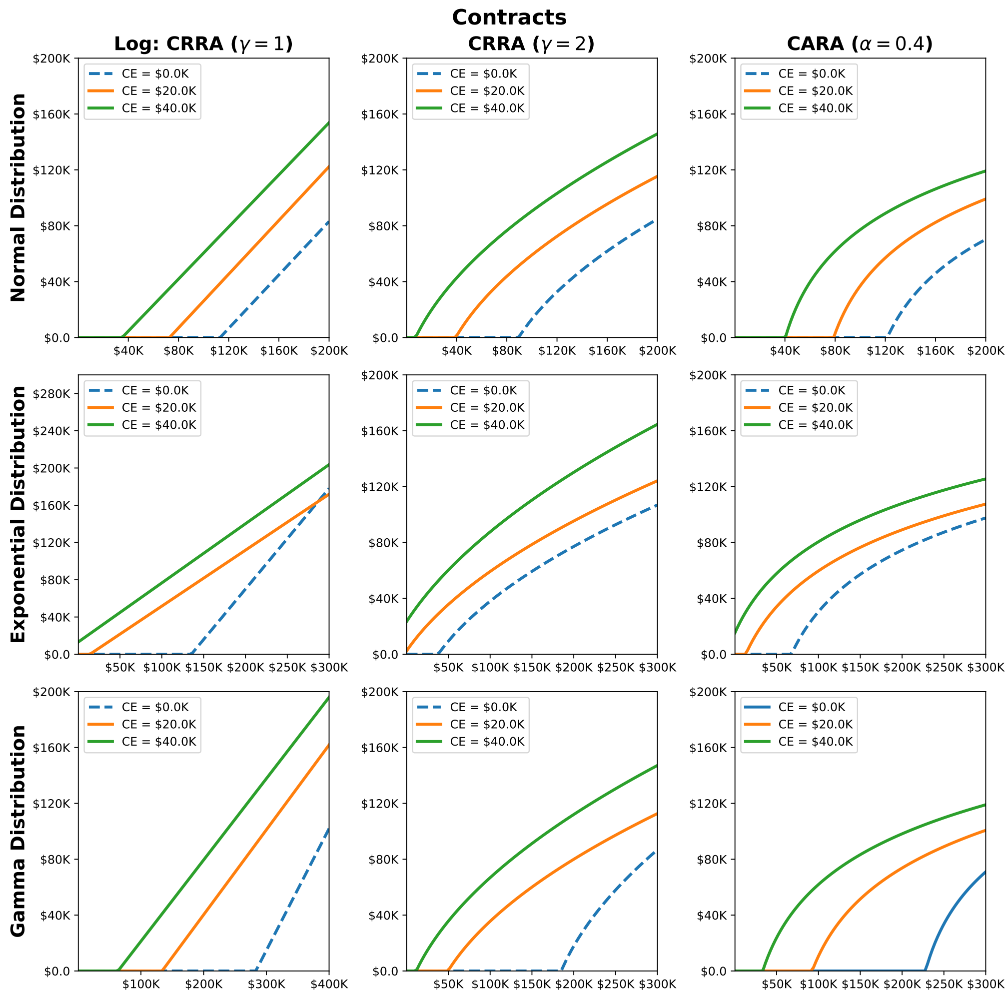
$$w(y) = k \circ g\left(\lambda + \mu S(y|a_0)\right). \tag{1}$$

The function $k \circ g$ is determined by the utility function and limited liability constraint. $S(y|a_0)$ is the score function (also known as the likelihood ratio), determined by the distribution of output. a_0 is the optimal action, and λ and μ are Lagrange multipliers.

Theorem 1 makes our limited liability setting particularly amenable to applications. As long as one considers a competitive setting with sufficiently high reservation utility,

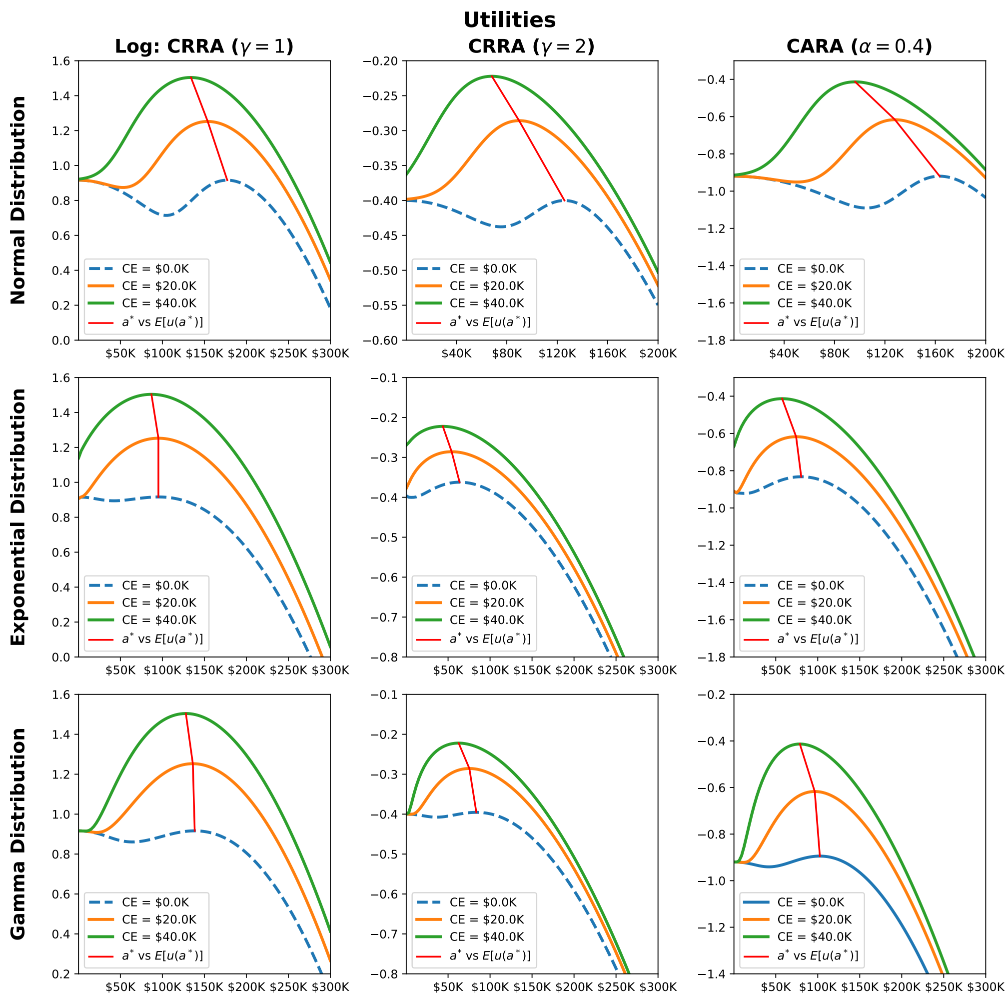
[5]These additional results are simple given the validity of the first-order approach. Our analysis and results are close to those in Jewitt et al. (2008), albeit we cannot simply use their results due to slightly different assumptions. For example, they assume the Rogerson (1985) conditions, which rule out examples such as the Gaussian distribution of output. Hence, we cannot simply use their results, although our proofs follow broadly similar lines.

Figure 2: Optimal Contracts under Different Distributions and Preferences



Note: This figure displays the wage function that solves the principal's profit maximization problem at different values of $\bar{U}$ for a grid of specifications. The log utility function is $u(x) = log(\frac{x}{20K})$. The CRRA utility function is $u(x) = \frac{1}{1-\gamma} \cdot \frac{x}{20K}^{1-\gamma}$, and $\gamma = 2$. The CARA utility function is $\frac{1}{\alpha} \cdot -\exp(-\alpha \frac{x}{20K})$, and $\alpha = \frac{2}{5}$. The normal distribution is $y \sim N(a, (20K^2))$. The exponential distribution's rate parameter is $\frac{1}{a}$. The gamma distribution has shape parameter $n = 2$, and scale parameter a. Formulas for all utility functions and output distributions are in the online appendix. The cost function is always $c(a) = \theta a^2$, but $\theta = .01$ in the log utility examples, $\theta = .004$ in the CRRA examples and $\theta = .008$ in the CARA examples. Our optimization routine has only one global IC constraint at $a = 1e - 3$. We conservatively say that the FOA is invalid if the Lagrange multiplier on the global IC constraint is greater than $1e - 5$ or if deviating to an action between 0 and .1 reduces the agent's utility by less than $1e - 2$.

Figure 3: Agent Utility versus Action under Different Distributions and Preferences



Note: This figure displays the agent's expected utility from a grid of possible actions given the corresponding wage functions from figure 2. The specifications are the same as in figure 2.

the main difficulties are resolved. Optimal contracts exist, are unique, and can be calculated trivially from equation (1).

Section 6 provides results to help researchers use the limited liability model. We provide simple calculus formulae for $k \circ g$ and S covering a wide set of examples. These can be readily applied in both theoretical and numerical models. We provide accompanying code for numerical solutions.

An immediate corollary of equation (1) is that piecewise linear contracts are optimal in a number of examples. Piecewise linear contracts are options, with a constant value up to a point, and increasing thereafter. The reason is the following. The function $k \circ g$ is piecewise linear when utility is log. And the score function is linear for many common statistical distributions of output, including Gaussian, exponential, Poisson, and Gamma. More broadly, in any distribution in an exponential family with linear sufficient statistic. This can be seen in figures (1) and (2). Optimal contracts are slightly nonlinear for very low reservation utilities, but piecewise linear as soon as the first-order approach starts holding. This extends previous results that rationalize option-like contracts (Innes, 1990; Jewitt et al., 2008; Chaigneau et al., 2024). This also contributes to the literature rationalizing simple contracts (Holmstrom and Milgrom, 1987; Carroll, 2015).

2.2 Relationship to the Literature

At first glance, our results appear to contradict the standard view in the literature that the first order approach requires stringent conditions. In fact, the two are perfectly consistent. Consider the Gaussian example in figure (1). When the reservation utility is low, the optimal contract produces two distinct local maxima of the agent's utility, violating the first-order approach because the agent then has a profitable non-local deviation. This is exactly the kind of global incentive problem highlighted by Mirrlees (1999). Likewise, the first-order approach fails in all the examples in figure (2). Hence, if one insists that the first-order approach must hold *for every reservation utility*, many interesting examples must be ruled out, matching the standard view in the literature.

Our contribution is to show that *if the reservation utility is sufficiently high*, those very same examples do satisfy the first-order approach. The core idea is that requiring the principal to give the agent greater overall utility elevates the payoff at the principal's intended action, pushing the local maximum at the intended action above other local maxima.

Historically, there are three generations of contributions to the first-order approach.

The first generation was started by Mirrlees (1999) in the mid 1970s, who showed that the first-order approach does not hold generally. Prior to Mirrlees (1999), the first-order approach was simply assumed with little justification. Subsequent papers then became aware that they had to assume validity of the first-order approach (Holmström, 1978), and developed results that do not depend on the first-order approach (Grossman and Hart, 1983).

The seminal papers in the second generation are Rogerson (1985) and Jewitt (1988). They provided sufficient conditions for the first-order approach. Rogerson's conditions are well known to be strict, ruling out many natural distributions, including the Gaussian and all examples in figure (2). Jewitt's conditions are more general, and include many interesting cases. Jewitt's main assumption is implicit, in that he requires his conditions to hold for all output levels, in particular ruling out limited liability (Chaigneau et al., 2022). Restrictions similar to limited liability are known to be important to guarantee existence (Moroni and Swinkels, 2014; Jewitt et al., 2008). In many examples that satisfy the Jewitt conditions, an optimal contract does not exist. This includes the Gaussian - log utility example without limited liability. The second-generation papers have been influential, and much of the theoretical and applied work to this day assumes these conditions.

Our own results are close to Jewitt (1988). We impose similar conditions on the part of the contract where limited liability does not bind, while allowing for limited liability. In contrast, Jewitt (1988) requires the conditions to hold globally, ruling out limited liability, and sometimes creating existence issues (Jewitt et al., 2008; Moroni and Swinkels, 2014). The proof of Theorem 1 shows that, with high reservation utility, the limited liability constraint binds with low probability. We then show, in similar lines to Jewitt (1988), that the agent's problem becomes concave.

What makes our result applicable is that, in many examples, the first-order approach starts holding for relatively modest reservation utility, and even in cases with multiple local maxima. This can be seen in figures (1) and (3). The intuition is that our theorem shows that, if reservation utility is very high, the $U(v^*, a)$ is concave in a. As reservation utility increases, the functions $U(v^*, a)$ have to unwind into a function with a single peak. And the recommended action becomes often much earlier than the function becomes concave. In this sense, our result has a limitation similar to other asymptotic results, such as the central limit theorem, that only holds for large enough sample sizes.

Finally, there is a third generation of work extending Jewitt's single-peakedness approach to multidimensional actions and richer information structures (e.g. Conlon, 2009;

Jung and Kim, 2015; Chade and Swinkels, 2020; Chaigneau et al., 2022; Jung et al., 2024). We believe our methods can be adapted to such environments, but we do not pursue that here.

3 Model

3.1 Model

A risk-neutral **principal** hires an **agent** with limited liability. The agent has utility $u(x) - c(a)$ of receiving a **payment** $x \in \mathbb{R}_+$ and taking **action** a in $\mathcal{A} \subseteq \mathbb{R}_+$. Both u and c are strictly increasing. **Output** $y \in \mathbb{R}$ depends on the agent's action and is distributed according to the **output density** $f(y|a)$.

A **contract** is a function $v : \mathbb{R} \to u(\mathbb{R}^+)$, which specifies the agent's utility $v(y)$ from the payment as a function of output y. Contracts are defined in terms of utility to simplify notation. To define the wage payments, let the **compensation cost function** k be the inverse of u. A **wage function** is a function $w : \mathbb{R} \to \mathbb{R}^+$. The wage function associated with contract v is $w(y) = k(v(y))$. Let $\mathcal{C}$ be the **set of feasible** contracts.

The **agent's utility from a contract** v given action a is

$$U(v,a) := \int v(y)f(y|a)\,dy - c(a).$$

The **expected wage** is

$$W(v,a) := \int k(v(y))f(y|a)\,dy.$$

The **cost minimization problem** given an **intended action** $a_0 \in \mathcal{A}$ and **reservation utility** $\bar{U}$ is to choose a contract v to minimize the expected wage subject to the individual rationality (IR) and global incentive compatibility (GIC) constraints. Individual rationality requires that the agent's expected utility is at least reservation utility, and global incentive compatibility requires that choosing a_0 is optimal for the agent. The cost minimization problem is to choose v in $\mathcal{C}$ to

$$\begin{aligned} \text{minimize} \quad & W(v) \\ \text{subject to} \quad & U(v, a_0) \geq \bar{U}, && \text{(IR)} \\ & U(v, a_0) \geq U(v, \hat{a}) \quad \forall \hat{a} \in \mathcal{A}. && \text{(GIC)} \end{aligned}$$

The **relaxed cost minimization problem** replaces the global incentive compatibility constraint with the local incentive compatibility constraint (LIC):

$$\begin{aligned} \text{minimize} \quad & W(v) \\ \text{subject to} \quad & U(v, a_0) \geq \bar{U}, && \text{(IR)} \\ & \partial_a U(v, a_0) = 0. && \text{(LIC)} \end{aligned}$$

3.2 Assumptions and Notation

We state here the assumptions needed for our results. This section can be skimmed on a first reading.

Assumption 1 (Regularity of utility, cost and density)**.** *The set of feasible actions* $\mathcal{A} \subseteq \mathbb{R}_+$ *is a compact interval including* 0*. The utility function* u *is strictly concave and smooth, and has* $\lim_{x \to \infty} u'(x) = 0$*. The cost function* c *is strictly convex and smooth, and* $c'(0) = 0$*. The output density* $f(y|a)$ *is smooth in the action,* a*, for all outputs,* y*.*

We now make technical assumptions that allow us to use Leibniz's rule of differentiation under the integral sign. For this, we impose restrictions on k, f and on the set of feasible contracts. At the same time, we require the set of feasible contracts to be rich enough so that relevant contracts are not ruled out by assumption.

Assumption 2 (Regularity of feasible contracts)**.** *The output density function* $f(y|a)$ *satisfies:*

1. *Every feasible contract is measurable.*

2. *Given any* a_1 *in* $\mathcal{A}$ *and a feasible contract* v*, there exists an integrable function* $\theta(y)$ *and a neighborhood of* a_1 *such that, for all* a *in this neighborhood,*

$$|v \partial_a^n f(y|a)| \leq \theta(y)$$

for $n = 0, 1, 2$ *and for all* y *in the support of* $f(\cdot|a)$*.*

3. *For all a in $\mathcal{A}$ and canonical contract v (as defined in definition 2),*

$$\Pi(v, a) < \infty.$$

4. *The set of feasible contracts is convex and includes all canonical contracts.*

We now consider substantive restrictions.

Assumption 3 (Assumptions on the distribution of output)**.** *The output density function $f(y|a)$ satisfies:*

1. *For all a, the support of $f(y|a)$ is $\mathbb{R}$.*

2. *$F_a(y|a) < 0$ for all y and a.*

3. *For all a, the score $\partial_a \log f(y|a)$ is strictly increasing in y and its image is $\mathbb{R}$ (this is known in the literature as the monotone likelihood ratio property or MLRP).*

4. *There exists y_0 in $\mathbb{R}$ such that, for all $y \leq y_0$ and a,*

$$f_{aa}(y|a) > 0.$$

Note that we assume that the support of f is $\mathbb{R}$. The numerical examples, like the exponential distribution, suggest that this assumptions is not necessary. We hope to weaken the assumption in the next version of this paper.

Assumption 4 (Concavity and limit of inverse marginal utility)**.** *The function k'^{-1} is strictly concave and*

$$\lim_{z \to \infty} z \frac{d}{dz} k'^{-1}(z)$$

is finite.

The function k'^{-1} is the same as Jewitt's (1988) $\omega(z)$. So our assumption that k'^{-1} is strictly concave corresponds to his assumption (2.12).

For some results, we restrict attention to intended actions where it is possible to satisfy the local incentive compatibility constraint.

Definition 1. *An intended action a is regular if it is in the interior of $\mathcal{A}$ and there exist a feasible contract v such that*

$$\partial_a U(v, a) \geq 0.$$

4 Main Result

We can now state our main result.

Theorem 1. *[Validity of the First Order Approach with High Reservation Utility] Given a regular intended action $a_0 > 0$, there exist reservation utilities U^* in $\mathbb{R}$ and $\bar{U}_R \in (U^*, \infty]$ such that:*

1. *For $\bar{U} \geq \bar{U}_R$, neither the relaxed cost miminization problem nor the cost minimization problem are feasible.*

2. *For $\bar{U} \leq \bar{U}_R$, the relaxed cost minimization problem has an almost everywhere unique solution given by proposition 1.*

3. *For $U^* \leq \bar{U} < \bar{U}_R$, this is also the almost everywhere unique solution of the cost minimization problem.*

5 Proof of the Main Result

The proof of the main result follows from two key propositions. Proposition 1 characterizes the solution of the relaxed problem. Proposition 2 shows that, for sufficiently high reservation utility, the agent's problem is concave. We now state these propositions and explain the key steps in the argument. Appendix A contains the proofs.

For the relaxed problem to be well-defined, the derivative $\partial_a U(v, a)$ must be well defined. Remark A.1 in the appendix shows that this is true under our assumptions, and moreover that we can calculate the derivative by differentiating under the integral sign. Henceforth, we will use differentiation under the integral sign without referencing remark A.1.

Throughout this section, fix a regular intended action a_0. To simplify notation, we omit the dependence on a_0 whenever it is clear, writing for example $U(v)$ instead of $U(v, a_0)$.

5.1 Solution to the Relaxed Problem

We first show that the relaxed cost minimization problem has an almost everywhere unique solution, and that this solution has a simple formula in terms of Lagrange multipliers. The form of the solution is based on the first order condition, and is similar to standard formulas in the literature, such as in the case without limited liability.[6]

The relaxed cost minimization problem is convex. Define its Lagrangian as

$$\mathcal{L}(v, \lambda, \mu) := W(v) + \lambda\left(\bar{U} - U(v)\right) + \mu(-U_a(v)). \tag{2}$$

Heuristically differentiating this Lagrangian with respect to $v(y)$ and setting the derivative to zero gives

$$k'(v(y))f(y|a) = \lambda f(y|a) + \mu f_a(y|a_0).$$

Dividing by $f(y|a)$ gives

$$k'(v(y)) = \lambda + \mu \frac{f_a(y|a_0)}{f(y|a_0)}. \tag{FOC}$$

Equation (FOC) is the key step in the standard first-order approach literature. Appendix A formally analyzes the convex program, demonstrates existence and uniqueness of the solution, and characterizes the optimal contract based on equation (FOC).

In our case of limited liability, the solution is described more clearly with the following notation. Define the **optimal expected wage** $\omega(\bar{U})$ as the value of the infimum in the relaxed cost minimization problem. Define the **link function** $g : \mathbb{R} \to \mathbb{R}$ as[7]

$$g(z) := k'^{-1}\left(\max\left\{\frac{1}{u'(0)}, z\right\}\right).$$

Define the **score** function[8] as in the statistics literature:

$$S(y|a) := \frac{\partial}{\partial a} \log f(y|a) = \frac{f_a(y|a)}{f(y|a)}.$$

---

[6]This type of formula for the optimal contract is well-known and central in the literature that assumes the first-order approach. The seminal reference is Holmström (1978), and recent examples with limited liability include Jewitt et al. (2008) and Chaigneau et al. (2022). The survey Georgiadis (2022) has an excellent explanation of the formula in its first paragraphs.

[7]The link function's input, z, is a marginal dollar cost of providing one util to the agent (measured in units of $\frac{\text{\$ of output}}{\text{utility}}$). The link function evaluated at z, $g(z)$, returns the utility level where z is the marginal cost of utility to the agent. Any z below $\frac{1}{u'(0)}$, which is the cheapest possible marginal cost, returns $g(z) = u(0)$.

[8]The score is known as the likelihood ratio in the economics literature. We favor the term score to be in line with the modern literature outside of economics.

Definition 2. *A canonical contract $V(y|\lambda,\mu)$ is defined for λ and μ in $\mathbb{R}$ as*

$$V\left(y|\lambda,\mu\right) := g\left(\lambda + \mu S(y|a_0)\right). \tag{3}$$

The following proposition shows that the relaxed problem typically has a solution and characterizes the solution and Pareto frontier.

Proposition 1. *[Solution of the Relaxed Cost-Minimization Problem] There exists $\bar{U}_L$ in $\mathbb{R}$ and $\bar{U}_R$ in $\mathbb{R}\cup\{\infty\}$ such that:*

1. *For $\bar{U} \geq \bar{U}_R$, the relaxed problem is not feasible.*
2. *For $\bar{U} < \bar{U}_R$, the relaxed problem has an almost everywhere unique solution $v^*(y|\bar{U})$. There exist Lagrange multipliers $\lambda^*(\bar{U}) \geq 0$ and $\mu^*(\bar{U}) > 0$ such that the solution is almost everywhere equal to the canonical contract*

$$v^*(y|\bar{U}) = V\left(y|\lambda^*(\bar{U}),\mu^*(\bar{U})\right).$$

3. *For $\bar{U} \leq \bar{U}_L$, we have $\lambda^*(\bar{U}) = 0$. The relaxed optimal contract $v^*(\cdot,\bar{U})$ and optimal expected wage $\omega(\bar{U})$ do not vary with $\bar{U}$ in this range.*
4. *For $\bar{U}_L < \bar{U} < \bar{U}_R$, $\lambda^*(\bar{U})$ is strictly increasing and $\lim_{\bar{U}\to\infty}\lambda^*(\bar{U}) = \infty$. The optimal expected wage $\omega(\bar{U})$ is strictly increasing and strictly convex.*

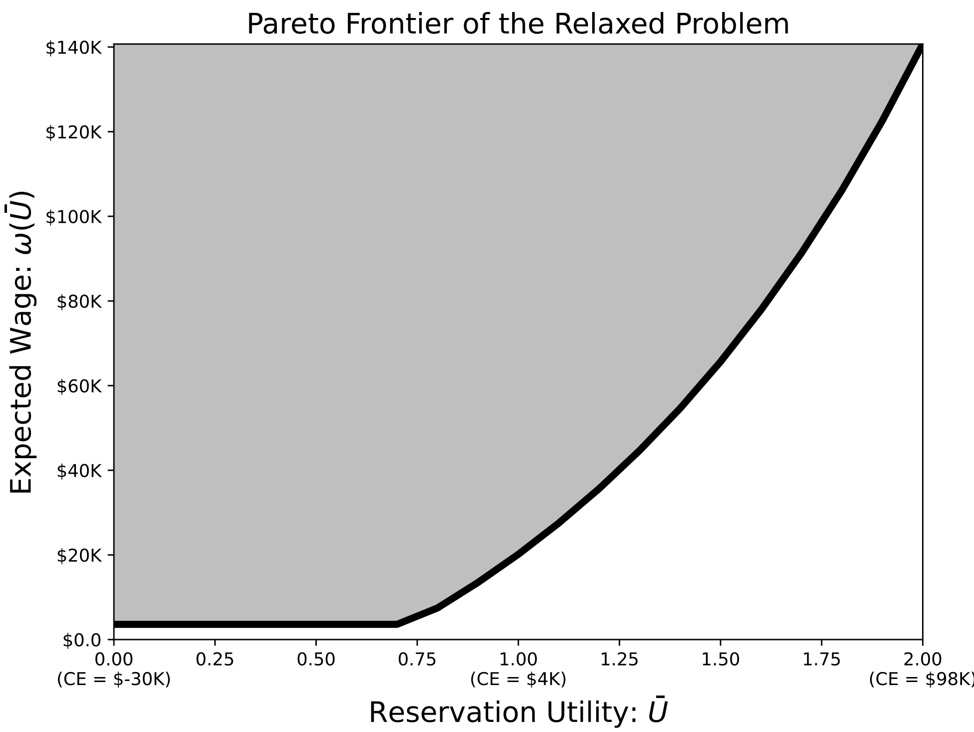
The Gaussian-log utility example illustrates proposition 1. The score is $(y-a)/\sigma^2$, the compensation cost function is $k(v) = \exp(v) - w_0$, and the link function is $\log\max\{z, w_0\}$. This implies the wage function

$$k(v^*(y|\bar{U})) = \left[\lambda^*(\bar{U}) + \mu^*(\bar{U})\frac{y-a}{\sigma^2} - w_0\right]^+.$$

This is the piecewise linear solution that we saw numerically in figure 1a. The convex Pareto frontier is illustrated in figure 4. In the log utility case, $\bar{U}_R = \infty$, as the principal can give arbitrarily high utility to the agent.

The formula in proposition 3 is essentially the standard first-order approach's equation (FOC), but properly accounting for limited liability. The intuition is the following. The principal would like to always pay the agent $g(\lambda^*(\bar{U}))$, which is the utility level where the marginal cost of providing utility to the agent is $\lambda^*(\bar{U})$. However, incentive

Figure 4: Pareto Frontier with Gaussian Distribution and Log Utility



Note: This figure displays the relaxed problem's Pareto frontier. The agent's utility is $u(x) = log(\frac{x}{20K})$, the cost function is $c(a) = \theta a^2$, where $\theta = .01$, and the agent's starting wealth is $w_0 = 50K$. The output is $y \sim N(a, (20K^2))$, where a is the agent's action. The intended action is $a_0 = 100K$.

compatibility requires that payment depends on whether there is statistical evidence of high effort. The contract, therefore, pays more if the score is positive and less if the score is negative.

Proposition 1 deviates substantially from the literature without limited liability. Our formula for the optimal contract is different because it gives the agent a constant payment of 0 for sufficiently low outcomes, where the limited liability constraint binds (see the kink in the wage function in Figure (5a). Proposition 1 guarantees existence of an optimum in the relaxed problem, which is impossible in a model without limited liability, where the relaxed problem often has no solution. This stark difference can be seen in the Gaussian-log utility example. Without limited liability, any contract satisfying the first-order conditions would pay a *globally* linear wage. But this would imply negative consumption in some states, and thus $-\infty$ utility to the agent. Indeed, the Gaussian-log utility example illustrates Mirrlees (1999)'s classic point that optimal contracts may not exist without limited liability.

5.2 The Case of High Reservation Utility

We now demonstrate that, for sufficiently high reservation utility, the solution to the relaxed cost minimization problem also solves the original cost minimization problem. To do so, we show that the relaxed problem's solution satisfies the global incentive compatibility constraint. This is achieved by demonstrating the stronger result that the agent's utility $U(v^*, a)$ is concave in a at the relaxed optimal contract.

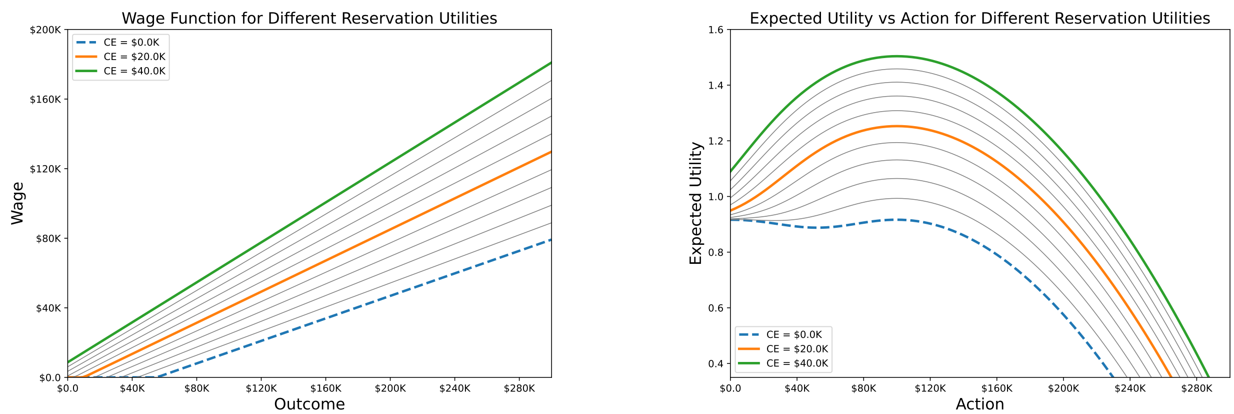
Proposition 2. *[Concavity of the Agent's Problem for High Reservation Utility]*
There exists $U^* < \bar{U}_R \in \mathbb{R}$ such that for all $\bar{U} > U^*$, and all $a \in \mathcal{A}$,

$$U_{aa}(v^*(y|\bar{U}), a) \leq 0.$$

The proposition shows that high reservation utility is crucial for the first order approach's validity, and the first order approach may not be valid at low reservation utility. Figure (5) helps illustrate this result by plotting the relaxed optimal contract, and the agent's expected utility versus her effort under the relaxed-optimal contract in the Gaussian-log utility example. The blue line represents the expected utility function for the relaxed-optimal contract with the lowest reservation utility, $\bar{U} = u(0)$. When reservation utility is low, the agent's problem is not concave, and choosing an action very close to

0 is a profitable global deviation for the agent. The limited liability constraint makes low effort appealing for the agent; it ensures that the agent can always achieve utility strictly greater than $u(0)$ by exerting no effort. If the IR constraint binds, as it does in many of our examples, then the agent's expected utility from choosing the intended action is $\bar{U}$, and the first order approach surely fails when $\bar{U} = u(0)$.

Figure 5: Relaxed Optimal Contracts with Gaussian Distribution and log Utility



(a) Canonical Contracts

(b) Expected Utility

Note: Panel a displays the wage function that solves the relaxed problem at different values of $\bar{U}$, and Panel b displays the agent's expected utility from a grid of possible actions given the corresponding wage functions from Panel a. The values of $\bar{U}$ are expressed in terms of certainty equivalents. The agent's utility is $u(x) = log(\frac{x}{20K})$, the cost function is $c(a) = \theta a^2$, where $\theta = .01$, and the agent's starting wealth is $w_0 = 50K$. The output is $y \sim N(a, (20K^2))$, where a is the agent's action. The intended action is $a_0 = 100K$.

Figure (5) also illustrates how, as reservation utility increases, the first-order approach becomes valid. The IR constraint ensures that $U(v^*, a_0)$ must be greater than $\bar{U}$. The rightmost local maximum, $U(v^*, a_0)$, therefore, increases as $\bar{U}$ increases. The leftmost local maximum also rises, but much more slowly. In the example, the first-order approach becomes valid with a very small increase in $\bar{U}$ above $u(0)$, as the rightmost local maximum overtakes the leftmost local maximum. Concavity of $U(v^*, a)$ on the other hand, usually requires much larger increases to $\bar{U}$, so, there are many values of $\bar{U}$ where the first-order approach is valid, even though $U(v^*, a)$ is not concave in a. Thus, concavity of $U(v, a)$ is a sufficient, but not necessary condition for the first-order approach's validity.

We now explain why the agent's problem is concave for high reservation utility, but not for low reservation utility. The relaxed optimal contract has two distinct regions. When the outcome, y, is below a threshold, the limited liability constraint binds, and the contract specifies a constant wage of 0. When y is above the threshold, the agent's payment is dictated by equation FOC. Figure 5 illustrates how the threshold depends on the agent's reservation utility. As reservation utility increases, the threshold shifts to the left.

The region where the limited liability constraint binds introduces convexity to the agent's problem. When the limited liability constraint binds, the agent is paid 0, and marginal increases to the agent's effort do not change her payment. The agent's returns to effort are therefore proportional to the probability the agent receives a positive payment. This probability is increasing on her effort, so the agent's expected wage is a convex function of the agent's effort.

Consider the Gaussian-log utility example. As we have seen, the wage function is

$$w(y) = \left(\lambda + \mu\frac{y - a_0}{\sigma^2} - w_0\right)^+ .$$

The outcome y can be decomposed into a normally distributed shock x plus the action a. Substituting into the wage function yields

$$w(y) = \left(\lambda + \mu\frac{x + a - a_0}{\sigma^2} - w_0\right)^+ .$$

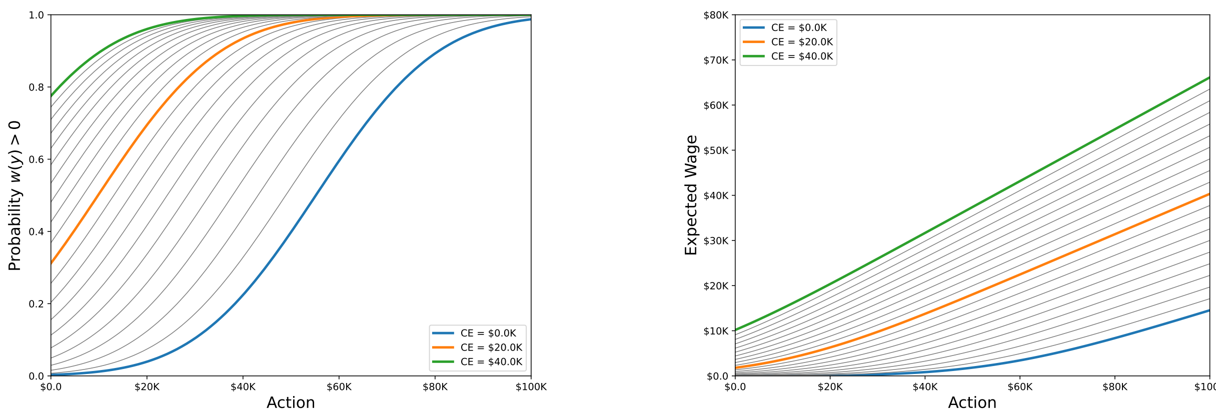
The derivative of the expected wage with respect to the action is

$$\frac{\partial \mathbb{E}[w(y)]}{\partial a} = \frac{\mu}{\sigma^2} \cdot P\left(\lambda + \mu\frac{x + a - a_0}{\sigma^2} - w_0 > 0\right).$$

The agent is only exposed to the outcome when x is sufficiently large. If x is too small for the agent to get paid, the agent receives no benefit from marginally increasing effort; she is paid 0 regardless. The agent experiences marginally increasing returns to effort because the probability that the agent's effort affects her payment increases on a.

As reservation utility increases, the threshold for positive payment shifts to the left, so the agent's probability of payment increases. Figure 6a illustrates that when reservation utility is high, the probability of payment is close to 1 even for low effort levels. Lemma A.7 formalizes the result that the probability of payment converges to 1 as reservation utility converges to $\bar{U}_R$. When reservation utility and the probability of payment are low, the probability of payment increases rapidly on effort, and the agent's expected wage function is highly convex (see Figure 6b). At high reservation utility, payment is likely regardless of the effort level, so the agent's expected wage is nearly linear. The agent's utility function is concave, and the cost function is convex, so the agent's problem is globally concave as long as the limited liability constraint does not distort the wage function too much.

Figure 6: Probability of Payment and Expected Wages in Gaussian-Log Utility Example



(a) Probability of Payment

(b) Expected Wage

Note: Panel A plots the probability that the agent receives a strictly positive wage against the agent's action, under the relaxed optimal contract for different values of reservation utility. Panel B plots the expected wage against effort for the same contracts. The parameters of the problem are as in Figure 5.

The proof of lemma A.7 illustrates why the probability of payment converges to 1 as reservation utility gets large. When reservation utility is large, the agent receives a large payment in nearly all states of the world where she is paid. If she were likely to be paid zero, there would be an incentive to work harder than the intended effort level to ensure that she receives the large payment. Therefore, for the first order condition to hold, the probability that the agent receives no payment must converge to 0.

Once this is established, proposition 2 follows naturally. Under our assumptions, the optimal contract v(y) is concave in the region where $w(y) > 0$. Thus, since the option-like region of the contract is far to the left, $U(v^*, a)$ is concave in a. Our result is closely related to Jewitt (1988)'s classic result. He shows that the agent's problem is concave in a similar setting, but without limited liability. In his setting, optimal contracts always have the concave shape that our contracts exhibit in the region $w(y) > 0$. Thus, Jewitt (1988) shows that the agent's problem is concave, and this holds for any reservation utility. Unfortunately, without limited liability there are often no optimal contracts, as we already demonstrated in the Gaussian-log utility example.

5.3 Proof of Theorem 1

Theorem 1 follows trivially from propositions 1 and 2.

Proof of Theorem 1. Parts (1) and (2) of Theorem 1 are included in Proposition 1. Part (3) follows from Proposition 2. For any $\bar{U} \geq U^*$ we have that $v^*(\bar{U})$ is concave, and thus is a solution to the cost minimization problem. Any other solution also solves the relaxed cost minimization problem, so equals $v^*(\bar{U})$ almost everywhere. □

6 Applications and Extensions

6.1 Closed Form Solutions

Theorem 1 implies that optimal contracts have simple functional forms in parametric settings. Here we describe the solutions for common parametrization and note their properties.

Optimal contracts depend on effort cost, risk preferences (which determine the link function g), and on the distribution of output (which determines the score S). Tables 1 and 2 provide basic formulae for the link and score functions that make up optimal contracts $g(\mu + \lambda S(y|a_0))$.

Table 1: Utility Functions, Link Functions, and Wage Functions

	Utility Function $u(x)$	Link Function $g(z)$	Wage Function $w(z)$
Log	$\log(x + w_0)$	$\log(\max(w_0, z))$	$(z - w_0)^+$
CRRA	$\frac{(x+w_0)^{1-\gamma}}{1-\gamma}$	$\frac{\max(w_0^\gamma, z)^{\frac{1-\gamma}{\gamma}}}{1-\gamma}$	$\left((z^+)^{\frac{1}{\gamma}} - w_0\right)^+$
CARA	$\frac{-\exp(-\alpha(x+w_0))}{\alpha}$	$-\frac{1}{\alpha \max(\exp(\alpha w_0), z)}$	$\frac{(\log^+ z - \alpha w_0)^+}{\alpha}$

Note: The utility function is the agent's utility from consumption given starting wealth w_0 and a transfer, x. The link and wage functions are in terms of z, which is a function of the outcome, y: $z(y) = \lambda + \mu S(y|a_0)$.

Linearity Linear contracts play a prominent role in contract theory. It has long been noted that linear contracts are common, but that they only arise under relatively special assumptions (Holmstrom and Milgrom, 1987; Carroll, 2015).

Table 2: Error Distributions

Distribution	Probability Density	Score Function $S(y\|a)$	Mean
Gaussian	$\mathcal{N}(a, \sigma^2)$	$\frac{y-a}{\sigma^2}$	a
Log Normal	$\frac{1}{y\sqrt{2\pi\sigma^2}} \exp\Big(-\frac{(\log(y)-a)^2}{2\sigma^2}\Big)$	$\frac{\log(y)-a}{\sigma^2}$	$\exp\Big(a+\frac{\sigma^2}{2}\Big)$
Poisson	$\frac{a^y e^{-a}}{y!}$	$\frac{y-a}{a}$	a
Exponential	$\frac{1}{a}e^{-\frac{y}{a}}$	$\frac{y-a}{a^2}$	a
Bernoulli	$a^y(1-a)^{1-y}, y \in \{0,1\}$	$\frac{y-a}{a-a^2}$	a
Geometric	$\Big(1-\frac{1}{a}\Big)^{y-1}\Big(\frac{1}{a}\Big), y \in \{1,2,\dots\}$	$\frac{y-a}{a^2-a}$	a
Binomial	$\binom{n}{y}a^y(1-a)^{n-y}, y \in \{0,\dots,n\}$	$\frac{y-na}{a-a^2}$	na
Gamma	$f(y \mid n,a) = \frac{y^{n-1}e^{-\frac{y}{a}}}{\Gamma(n)\,a^n}$	$\frac{y-na}{a^2}$	na
Student's t	$\frac{\Gamma\big(\frac{\nu+1}{2}\big)}{\Gamma\big(\frac{\nu}{2}\big)\sqrt{\pi\nu}\,\sigma}\Big(1+\frac{1}{\nu}\frac{(y-a)^2}{\sigma^2}\Big)^{-\frac{\nu+1}{2}}$	$\frac{(\nu+1)(y-a)}{\nu\sigma^2+(y-a)^2}$	a
Exponential Family	$h(y)\,\exp\Big(\eta(a)\,T(y)-A(a)\Big)$	$T(y)\,\frac{d\eta(a)}{da} - \frac{dA(a)}{da}$	*(Not specified)*
$y = a + X, X \sim h$	$g(y-a)$	$-\frac{g'(y-a)}{h(y-a)}$	$a + \mathbb{E}[X]$
$y = aX, X \sim h$	$\big\|\frac{1}{a}\big\|\, h\Big(\frac{y}{a}\Big)$	$-\frac{1}{a} - \frac{y}{a^2}\frac{g'(\frac{y}{a})}{g(\frac{y}{a})}$	$a\,\mathbb{E}[X]$

Note: This table presents probability the PDF, score function, and means of probability distributions as functions of the agent's chosen action, a.

Our model yields piecewise linear contracts in a number of examples, as seen in the tables and in the Gaussian-log utility example. The key ingredients are log utility, which makes the wage function linear in the score, and a linear score function. This includes many distributions, because the score is linear for distributions in an exponential family with linear sufficient statistics. That is, when $f(y, a)$ is of the form

$$f(y|a) = \exp\left(\eta(a)y + A(a)\right). \tag{4}$$

We note this as follows:

Remark 1 (Linear Contracts)**.** *Assume that utility is log ($u(x) = \log(w_0 + x)$), and that the distribution of outcomes is in an exponential family with linear sufficient statistic (as in equation 4). Then the relaxed optimal contract wage function is piecewise linear. This includes the Gaussian, exponential, Poisson, and gamma distributions.*[9]

[9] Note that the remark covers distributions with support different than $\mathbb{R}$, such as the exponential. These distributions do not satisfy Assumption (3). Nevertheless, the proof of proposition 1 does not use that the

Concave and Convex Contracts The optimal contract is always at least partially convex because the limited liability constraint requires that $w(y) = 0$ for all y less than some threshold, $\underline{y}$. The region where the limited liability constraint does not bind, however, can be convex or concave depending on the agent's risk aversion.

Remark 2. *Suppose the score function is linear, and let $\mu S(y|a_0) = ky$. Then, if the agent has CRRA utility, the region of the optimal contract where the limited liability constraint does not bind is convex if the risk aversion parameter $\gamma > 1$ and concave if $\gamma < 1$:*[10]

$$w(y) = (\lambda + ky)^{\frac{1}{\gamma}} \textit{ for } y > \underline{y}.$$

If the agent has CARA utility, the same region of the optimal contract is logarithmic:

$$w(y) = \frac{\log(\lambda + ky)}{\alpha} - w_0 \textit{ for } y > \underline{y}.$$

The remark shows that the shape of the optimal contract depends heavily on the agent's utility function, and a wide variety of contracts are potentially compatible with the theory. Our result is contrary to Conlon (2009)'s view that convex contracts, like CEO compensation with stock options, are incompatible with the first-order approach. In fact, the optimal contract with CRRA utility and $\gamma < 1$ closely approximates a CEO compensation package of stock options with many strike prices. To see this, consider a CEO who receives n_i options with strike price s_i that expire at the end of the year. The CEO's wage is

$$w(y) = \sum_i n_i (y - strike_i)^+,$$

where y is the stock's end of year price price. The wage function's slope at y is $\sum_{strike_i < y} n_i$; the wage is a piecewise-linear function where the slope increases as y increases. The options package wage may be a discretization of the contract the model predicts for an agent with CRRA utility with $\gamma < 1$, and a linear score function. Its slope is also increasing on the outcome and equals $\frac{k}{\gamma}(\lambda + ky)^{\frac{1-\gamma}{\gamma}}$ for some constant k.

Remark 2 affirms Holmstrom and Milgrom's (1987) view that the principle agent

support is $\mathbb{R}$, so the remark holds. However, our proof that the relaxed optimal contract is optimal does use that the support is $\mathbb{R}$. Therefore, we do not know whether Theorem 1 can be extended to different supports, and thus whether the remark can be extended to optimal contracts in the case of sufficiently high reservation utility. We conjecture based on numerical results that this is true and hope to include these facts in our next revision.

[10] Assumption 4 requires that $\gamma > \frac{1}{2}$.

model does not robustly predict linear wages. The optimal wage is only linear if the agent's risk aversion parameter is precisely $\gamma = 1$. However, if the agent's risk aversion is approximately one, the optimal contract will be approximately linear, and using a linear contract may be practical.

Limitations and numerical methods. The formulae have two main limitations. First, there is no general closed form solution for the Lagrange multipliers λ and μ. Thus, Theorem 1 guarantees that optimal contracts have the described functional form, but the Lagrange multipliers have to be calculated numerically. Second, as noted in Theorem 1, these simple formulae do not hold for sufficiently low reservation utility.

To address these limitations, we provide accompanying code to numerically solve for optimal contracts. The numerical methods solve both the relaxed problem, and also the full cost minimization problem, in the case where only a finite number of global constraints bind. This is the case in all the experiments that we conducted. Both are convex optimization problems, and can be solved at trivial computational cost in all our experiments. The code implements all the common specifications in tables 1 and 2.

7 Conclusion

This paper shows that when an agent's reservation utility is sufficiently high, the long-standing concerns about the validity of the first-order approach largely disappear. Despite the possibility of multiple local maxima under low reservation utility, forcing the principal to provide higher overall utility eliminates global deviations, ensuring both existence of an optimal contract and applicability of the standard calculus-based approach. Our results hold under natural assumptions and extend to many commonly used distributions and preferences. This suggests that the first-order approach is especially relevant in competitive labor markets or other environments where agents have strong outside options.

In particular, our results show that the many existing results in the literature that have been proven assuming the first-order approach hold in this broader array of settings, include many settings with option-like contracts, and even contracts with wages that are convex functions of output. We hope that this result and the basic calculus formulae we provide will be useful in applied and theoretical work.

References

Bolton, Patrick and Mathias Dewatripont, *Contract Theory*, Cambridge, MA: MIT Press, 2005.

Carroll, Gabriel, "Robustness and linear contracts," *American Economic Review*, 2015, *105* (2), 536–563.

Chade, Hector and Jeroen Swinkels, "The no-upward-crossing condition, comparative statics, and the moral-hazard problem," *Theoretical Economics*, 2020, *15* (2), 445–476.

Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb, "How should performance signals affect contracts?," *The Review of Financial Studies*, 2022, *35* (1), 168–206.

_ , _ , and _ , "A Theory of Fair CEO Pay," *Corporate Finance*, 2024, *3*, 4.

Conlon, John R, "Two new conditions supporting the first-order approach to multisignal principal–agent problems," *Econometrica*, 2009, *77* (1), 249–278.

Edmans, Alex, Xavier Gabaix, and Augustin Landier, "A multiplicative model of optimal CEO incentives in market equilibrium," *The Review of Financial Studies*, 2009, *22* (12), 4881–4917.

Georgiadis, George, "Contracting with moral hazard: A review of theory & empirics," *Available at SSRN 4196247*, 2022.

Grossman, Sanford J. and Oliver D. Hart, "An Analysis of the Principal-Agent Problem," *Econometrica*, 1983, *51* (1), 7–45.

Holmström, Bengt, "Moral hazard and observability," *The Bell journal of economics*, 1979, pp. 74–91.

Holmstrom, Bengt and Paul Milgrom, "Aggregation and linearity in the provision of intertemporal incentives," *Econometrica: Journal of the Econometric Society*, 1987, pp. 303–328.

Holmström, Bengt Robert, "On Incentives and Control in Organizations." PhD dissertation 1978.

Innes, Robert D, "Limited liability and incentive contracting with ex-ante action choices," *Journal of economic theory*, 1990, *52* (1), 45–67.

Jewitt, Ian, "Justifying the first-order approach to principal-agent problems," *Econometrica: Journal of the Econometric Society*, 1988, pp. 1177–1190.

— **, Ohad Kadan, and Jeroen M Swinkels**, "Moral hazard with bounded payments," *Journal of Economic Theory*, 2008, *143* (1), 59–82.

Jung, Jin Yong and Son Ku Kim, "Information space conditions for the first-order approach in agency problems," *Journal of Economic Theory*, 2015, *160*, 243–279.

— **, Son-Ku Kim, and Seung Joo Lee**, "A Proxy Contract Based Approach to the First-Order Approach in Agency Models," *Available at SSRN 4899689*, 2024.

Kadan, Ohad, Philip J Reny, and Jeroen M Swinkels, "Existence of optimal mechanisms in principal-agent problems," *Econometrica*, 2017, *85* (3), 769–823.

Mirrlees, James A, "The Theory of Moral Hazard and Unobservable Behaviour: Part I," *The Review of Economic Studies*, 1999, *66* (1), 3–21.

Moroni, Sofia and Jeroen Swinkels, "Existence and non-existence in the moral hazard problem," *Journal of Economic Theory*, 2014, *150*, 668–682.

Rogerson, William P., "The First-Order Approach to Principal-Agent Problems," *Econometrica*, 1985, *53* (6), 1357–1367.

Salanié, Bernard, *The economics of contracts: a primer*, MIT press, 2005.

Sinclair-Desgagné, Bernard, "The first-order approach to multi-signal principal-agent problems," *Econometrica: Journal of the Econometric Society*, 1994, pp. 459–465.

Zeckhauser, Richard, "Medical insurance: A case study of the tradeoff between risk spreading and appropriate incentives," *Journal of Economic theory*, 1970, *2* (1), 10–26.

Appendix

A Proofs

A.1 Solution of the Relaxed Problem

Throughout this subsection, fix a reservation utility $\bar{U}$ and regular intended action a_0. The relaxed problem only has a solution if $U(v, a)$ can be differentiated under the integral sign, so we begin by showing that differentiation under the integral at any action is possible for any feasible contract.

Remark A.1. *Let v be a feasible contract. Then, for all a in $\mathbb{R}_+$, $U_a(v, a)$ exists, is finite, and*

$$U_a(v, a) = \int v(y) f_a(y|a)\, dy.$$

Proof. The conditions for Leibniz's rule to hold are in Assumption 2.2. □

We now derive some intermediate results to establish proposition 1. Recall the definition of the Lagrangian from equation (2) and definition 2 of a canonical contract.

Note that canonical contracts uniquely minimize the Lagrangian:

Lemma A.1. *Given λ and μ in $\mathbb{R}$, there exists v that minimizes the Lagrangian $\mathcal{L}(v, \lambda, \mu)$ among all feasible contracts. v is $f(\cdot|a_0)$ almost everywhere equal to the canonical contract $V(y, \lambda, \mu)$.*

Proof. The Lagrangian in equation (2) can be written as

$$\mathcal{L}(v, \lambda, \mu) = \int [k(v(y)) - \lambda v(y) - \mu v(y) S(y|a_0)] f(y|a_0)\, dy + \bar{U}.$$

Differentiating the integrand with respect to $v(y)$ yields

$$[k'(v(y)) - \lambda - \mu S(y|a_0)] f(y|a_0)$$

and this is strictly convex in $v(y)$. Therefore, the integrand is minimized pointwise in $v(y)$ at $v(y) = V(y|\lambda, \mu)$. Hence, the infimum is attained, and any minimizer satisfies the desired formula $f(y|a_0)$ almost everywhere.

It only remains to show that $V(y|\lambda, \mu)$ is feasible. This follows from assumption 2 Part 4. □

We now note that, given λ there is a unique value of μ that solves the local IC constraint.

Lemma A.2. *Given λ in $\mathbb{R}$, there exists a unique $\tilde{\mu}(\lambda)$ such that the canonical contract $V(y|\lambda, \tilde{\mu}(\lambda))$ satisfies the local IC constraint (LIC). Moreover, $\tilde{\mu}(\lambda) > 0$.*

Proof. We have

$$U_a(v) = \int v(y) f_a(y|a_0)\, dy.$$

If $\mu = 0$, then $v(y)$ is constant, so

$$U_a(V(\cdot|\lambda, 0, a_0)) - c'(a_0) < 0.$$

As $\mu \to \infty$, $V(y|\lambda, \mu, a_0)$ converges pointwise to $u(\infty)$ if $S(y|a_0) > 0$ and to $u(0)$ if $S(y|a_0) < 0$). Hence, because a_0 is regular, for large enough μ,

$$U_a(V(\cdot|\lambda, \mu, a_0)) - c'(a_0) > 0.$$

Therefore, there exists at least one $\mu_1 > 0$ such that

$$U_a(V(\cdot|\lambda, \mu_1, a_0)) - c'(a_0) = 0.$$

It only remains to prove that this solution μ_1 is unique. To see this, note that

$$U_a(V(\cdot|\lambda, \mu_1, a_0))$$

is weakly increasing in μ. And, moreover, it is strictly increasing at any solution μ_1 because $c'(a_0) > 0$ implies that there is a positive measure of y such that $f_a(y|a_0) > 0$ and

$$\lambda + \mu_1 \frac{f_a(y|a_0)}{f(y|a_0)} > \frac{1}{u'(0)}.$$

□

The lemma implies that the family of canonical contracts that satisfy (LIC) is a one-dimensional family indexed by λ. Define

$$\tilde{V}(y|\lambda) := V(y|\lambda, \tilde{\mu}(\lambda))$$

Define the **relaxed Pareto problem** as finding v in $\mathcal{C}$ to

$$\begin{aligned} \text{minimize} \quad & W(v) - \lambda U(v) \\ \text{subject to} \quad & \partial_a U(v, a_0) = 0 \end{aligned} \tag{LIC}$$

The next lemma shows that the contracts $\tilde{V}(y|\lambda)$ span the solutions to the Pareto problem:

Lemma A.3. *Given λ in $\mathbb{R}$, the relaxed Pareto problem has a solution, and any solution is $f(y|a_0)$ almost everywhere equal to $\tilde{V}(y|\lambda)$.*

Proof. For any v satisfying (LIC),

$$W(v) - \lambda U(v) = \mathcal{L}(v, \lambda, \tilde{\mu}(\lambda)) - \lambda \bar{U}.$$

Lemma A.1 then implies that $W(v) - \lambda(v)$ is minimized over $\mathcal{C}$ by $\tilde{V}(y|\lambda)$, and that this solution is almost-everywhere unique. This contract satisfies (LIC) by the definition of $\tilde{\mu}(\lambda)$. □

Define the expected wage and utility attained by these contracts as

$$\begin{aligned} \tilde{U}(\lambda) \quad &:= \quad U(\tilde{V}(\cdot|\lambda), a_0), \\ \tilde{W}(\lambda) \quad &:= \quad W(\tilde{V}(\cdot|\lambda), a_0). \end{aligned}$$

Lemma A.4. *$\tilde{U}(\lambda)$ is strictly increasing.*

Proof. Consider $\lambda_1 < \lambda_2$ with optima $v_1 := \tilde{V}(\cdot|\lambda_1)$ and $v_2 := \tilde{V}(\cdot|\lambda_2)$. By optimality,

$$W(v_1) - \lambda_1 U(v_1) \leq W(v_2) - \lambda_1 U(v_2),$$

and

$$W(v_2) - \lambda_2 U(v_2) \leq W(v_1) - \lambda_2 U(v_1).$$

Adding the inequalities,

$$(\lambda_2 - \lambda_1) \cdot \left(\tilde{U}(\lambda_2) - \tilde{U}(\lambda_1)\right) \geq 0.$$

Therefore, $\tilde{U}$ is non-decreasing. It only remains to show that $\tilde{U}$ is strictly increasing. To reach a contradiction, assume that $\tilde{U}(\lambda_2) = \tilde{U}(\lambda_1)$. Optimality implies that $\tilde{W}(\lambda_1) =$

$\tilde{W}(\lambda_2)$. Let $\bar{\lambda} = (\lambda_1 + \lambda_2)/2$. We have $\tilde{U}(\bar{\lambda}) = \tilde{U}(\lambda_1)$. By our assumptions on the score, v_1 and v_2 differ in a set of positive measure. By strict convexity of k, it follows that $\tilde{W}(\bar{\lambda}) < \tilde{W}(\lambda_1)$. This contradicts the optimality of v_1. □

The proof of proposition 1 follows from collecting these results.

Proof of Proposition 1. Let $\bar{U}_L := \tilde{U}(0)$ and $\bar{U}_R := \lim_{\lambda\to\infty} \tilde{U}(\lambda)$.

Part 1.

$\bar{U}_R$ equals $u(\infty) - c(a_0)$, so the relaxed problem is not feasible for $\bar{U} \geq \bar{U}_R$, as desired.

Part 2.

Let $\lambda^*(\bar{U})$ be 0 if $\bar{U} \leq \bar{U}_L$ and be the inverse of $\tilde{U}$ for $\bar{U}_L < \bar{U} < \bar{U}_R$. Lemma A.4 implies that λ^* is well defined and that $\lambda^*(\bar{U}) \geq 0$.

Let $\mu^*(\lambda) := \tilde{\mu}(\lambda^*(\bar{U}))$. Lemma A.2 implies that $\mu^*(\bar{U}) > 0$. Let $v^*(y|\bar{U}) := \tilde{V}(y|\lambda^*(\bar{U}))$. Note that this coincides with the definition of v^* in the proposition statement.

Lemma A.3 implies that $v^*(y|\bar{U})$ solves the Pareto problem given $\lambda^*(\bar{U})$. This implies that $v^*(y|\bar{U})$ also solves the relaxed cost minimization problem. Likewise, A.3 implies that the solution is unique almost everywhere.

Part 3.

This follows from the definition of $\lambda^*(\bar{U})$ which equals 0 in this range.

Part 4. λ^* strictly increasing follows from lemma *A*.4 and the strict convexity of ω follows from $v^*(\cdot|\lambda)$ being different for each value of λ and from k being strictly convex.

□

A.2 Proof of Proposition 2

This section demonstrates that the second derivative of agent's utility is negative for sufficiently high reservation utility. The bulk of the section is spent demonstrating the same result for λ sufficiently high, and the main result follows as an immediate corollary. Throughout this section, we use the $\tilde{V}(y|\lambda)$ notation because we are focused on the agent's problem as λ becomes large. Although μ is a function of λ (Lemma A.2) we abuse notation by omitting μ's dependence on λ and using $\tilde{\mu}$ to denote $\tilde{\mu}(\lambda)$.

The section proceeds as follows. Lemmas A.5 and A.6 derive a convenient formula for the second derivative of agent's utility with respect to effort. Lemma A.7 shows that the agent's probability of receiving payment approaches 1 as λ approaches infinity. Lemma A.8 shows that the second derivative of agent's utility is negative for λ sufficiently high.

We then use these lemmas to prove the main result. Finally, we prove a few minor lemmas which we used in the proofs of the major lemmas.

We begin with an equation for the general form of agent's utility and its derivatives for an arbitrary contract.

Lemma A.5. *Given a contract, v, U and its derivatives evaluated at a equal*

$$\begin{aligned} U(v,a) &= \int v \cdot f(v|a)\, dy - c(a), \\ U_a(v,a) &= \int v \cdot S(y|a) f(y|a)\, dy - c'(a), \\ U_{aa}(v,a) &= \int v \cdot \left(S^2(y|a) + S_a(y|a)\right) f(v|a)\, dy - c''(a). \end{aligned}$$

Proof. The expression for $U(v,a)$ is just the definition. Differentiating $U(v,a)$ with respect to a yields

$$U_a(v,a) = \int v f_a(v|a)\, dv - c'(a).$$

The formula for U_a follows from the fact that $f_a(v|a) = f(v|a)S(v|a)$. Differentiating $U_a(v,a)$ with respect to a gives

$$\begin{aligned} U_{aa}(v,a) &= \int v \frac{\partial}{\partial a} \left[f(v|a) S(v|a) \right] dv - c'(a) \\ &= \int v \left[f_a(v|a) S(v|a) + f(v|a) S_a(v|a) \right] dv - c''(a) \\ &= \int v \left[f(v|a) S(v|a) S(v|a) + f(v|a) S_a(v|a) \right] dv - c''(a) \\ &= \int v \left[S^2(v|a) + S_a(v|a) \right] f(v|a)\, dv - c''(a). \end{aligned}$$

□

We now use Lemma A.5 to derive the following equations for the agent's utility and its derivatives.

Lemma A.6. *Given a canonical contract, $v(y) := \tilde{V}(y|\lambda)$, the agent's utility and its derivatives* *evaluated at a are*

$$U(v,a) = g(\lambda) + \tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a_0) f(y|a)\, dy - c(a),$$
$$U_a(v,a) = \tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a) \cdot S(y|a_0) f(y|a)\, dy - c'(a),$$
$$U_{aa}(v,a) = \tilde{\mu} \int \Delta g(y|\lambda) \left(S^2(y|a) + S_a(y|a) \right) \cdot S(y|a_0) f(y|a)\, dy - c''(a),$$

where

$$\Delta g(y|\lambda) = \frac{g(\lambda + \tilde{\mu} S(y|a_0)) - g(\lambda)}{\tilde{\mu} S(y|a_0)}.$$

Proof. The utility function, U, evaluated at the canonical contract, $\tilde{V}(y|\lambda)$, and action, a, is

$$U\left(\tilde{V}(y|\lambda), a\right) = \int \tilde{V}(y|\lambda) f(y|a)\, dy.$$

Recall that $\tilde{V}(y|\lambda) = g\left(\lambda + \tilde{\mu} S(y|a_0)\right)$. Substituting in yields

$$U\left(\tilde{V}(y|\lambda), a\right) = \int g\left(\lambda + \tilde{\mu} S(y|a_0)\right) f(y|a)\, dy.$$

Adding and subtracting $g(\lambda)$, we rewrite the integral:

$$U\left(\tilde{V}(y|\lambda), a\right) = g(\lambda) + \int \left[g\left(\lambda + \tilde{\mu} S(y|a_0)\right) - g(\lambda)\right] f(y|a)\, dy.$$

Using the definition $\Delta g(y|\lambda) = \frac{g(\lambda + \tilde{\mu} S(y|a_0)) - g(\lambda)}{\tilde{\mu} S(y|a_0)}$ and multiplying by $\frac{\tilde{\mu} S(y|a_0)}{\tilde{\mu} S(y|a_0)}$ yields

$$U\left(\tilde{V}(y|\lambda), a\right) = g(\lambda) + \tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a_0) f(y|a)\, dy - c(a).$$

The lemma follows from A.5's formula for the derivatives applied to $v(y) = \Delta g(y|\lambda) \cdot S(y|a_0)$. □

We now demonstrate that for λ sufficiently large the agent receives a strictly positive payment for an arbitrarily large portion of the support of $f(y|a)$ for any $a > 0$. We begin by defining some additional notation.

Definition A.1. *The* ***threshold score****,* $\underline{S}(\lambda)$*, is the maximum score such that the agent receives no payment. It is the score that solves*

$$\underline{S}(\lambda) = \frac{1}{\tilde{\mu} u'(0)} - \frac{\lambda}{\tilde{\mu}}.$$

The threshold score is well defined because $\tilde{\mu} > 0$ by Lemma A.2.

Definition A.2. *The **threshold outcome**, $\underline{y}(\lambda)$, is the outcome which induces the threshold score, $\underline{S}(\lambda)$:*

$$\underline{y}(\lambda) = S^{-1}(\underline{S}(\lambda, \tilde{\mu})|a_0).$$

A score that satisfies the equation exists by Assumption 3.3, which states that the score's image is $\mathbb{R}$.

Lemma A.7. *The threshold outcome approaches negative infinity as λ approaches infinity:*

$$\lim_{\lambda \to \infty} \underline{y}(\lambda) = -\infty.$$

Proof. Lemma A.6's result for U_a evaluated at a_0 yields

$$U_a\left(\tilde{V}(y|\lambda), a_0\right) = \tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a_0)^2 f(y|a_0) - c'(a_0).$$

The local incentive compatibility constraint requires that $U_a\left(\tilde{V}(y|\lambda), a_0\right) = 0$. It follows that

$$\tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a_0)^2 f(y|a_0) = c'(a_0).$$

Observe that the term inside the expectation is weakly positive because $\Delta g(y|\lambda) \geq 0$ by the monotonicity of g and $S(y|a_0)^2 \geq 0$. Therefore,

$$c'(a_0) \geq \tilde{\mu} \int_{S(y|a_0) \leq \underline{S}(\lambda)} \Delta g(y|\lambda) \cdot S^2(y|a_0) f(y|a_0)\, dy.$$

By definition, for all y in the domain of integration, $g(y) = u(0)$. Substituting in to $\Delta g(y|\lambda)$ yields

$$c'(a_0) \geq \tilde{\mu} \int_{S(y|a_0) \leq \underline{S}(\lambda)} \frac{(u(0) - g(\lambda))}{\tilde{\mu} S(y|a_0)} \cdot S^2(y|a_0) f(y|a_0)\, dy.$$

We simplify and use the fact that $f_a(y|a_0) = f(y|a_0)S(y|a_0)$

$$c'(a_0) \geq (u(0) - g(\lambda)) \cdot \int_{S(y|a_0) \leq \underline{S}(\lambda)} f_a(y|a_0)\, dy.$$

Integrating yields:

$$c'(a_0) \geq (u(0) - g(\lambda)) \cdot F_a(\underline{y}(\lambda)|a_0).$$

Observe that $u(0) - g(\lambda) \to -\infty$, and $F_a(\underline{y}(\lambda)|a_0) < 0$ by Assumption 3.2. Suppose $F_a(\underline{y}(\lambda)|a_0)$ does not approach 0. Then we have $c'(a_0) \geq \infty$. By contradiction, $F_a(\underline{y}(\lambda)|a_0) \to 0$.

The threshold outcome, $\underline{y}(\lambda)$, cannot converge to infinity because $\underline{S}(\lambda)$ is negative for λ sufficiently large because $S(\infty|a_0) = \infty$ by Assumption 3.3, and $\infty > \underline{S}(\lambda)$. If $\underline{S}(\lambda)$ does not converge to negative infinity, then there is a subsequence that converges to a finite number. However, $F_a(c|a_0) < 0$ for any c by Assumption 3.2. The proposition is proven by contradiction. □

Lemma A.8. *As λ approaches infinity, the limit of the supremum of the second derivative of agent's utility at any action $a > 0$ is strictly negative:*

$$\limsup_{\lambda \to \infty} U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) < 0.$$

Proof. By Lemma A.6,

$$U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) = g(\lambda) + \tilde{\mu} \int \Delta g(y|\lambda) \cdot S(y|a_0) f(y|a)\, dy - c(a)$$

Lemma A.9 states that there exists a y_0 such that the integrand is negative for all $y < y_0$. It follows that

$$U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) \leq \tilde{\mu} \int_{y_0}^{\infty} \Delta g(y|\lambda) \left(S^2(y|a) + S_a(y|a)\right) \cdot S(y|a_0) \cdot f(y|\hat{a})\, dy - c''(a).$$

Let Y_+ be the set of $y \geq y_0$ where this integrand is positive. Then

$$U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) \leq \tilde{\mu} \int_{Y_+} \Delta g(y|\lambda) \left(S^2(y|a) + S_a(y|a)\right) \cdot S(y|a_0) \cdot f(y|\hat{a})\, dy - c''(a).$$

Lemma A.7 implies that for λ sufficiently large, $\underline{y}(\lambda, a_0) \leq y_0$. Recall that for $y > \underline{y}$, $g(y) = k'^{-1}(\lambda + \mu S(y|a_0))$. The concavity of k'^{-1} (Assumption 4) implies that Δg is decreasing for y in Y_+. Therefore,

$$U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) \leq \tilde{\mu} \Delta g(y_0|\lambda) \int_{Y_+} \left(S^2(y|a) + S_a(y|a)\right) \cdot S(y_0|a_0) \cdot f(y|\hat{a})\, dy - c''(a).$$

Because $g(y)$ is concave for $y > \underline{y}$ and $y_0 > \underline{y}$

$$
\begin{aligned}
\tilde{\mu}\Delta g(y_0|\lambda) &\leq \tilde{\mu} g'(\lambda + \tilde{\mu} S(y|a_0)) \\
&= \frac{\tilde{\mu}}{\lambda + \tilde{\mu} S(y|a_0)} \cdot (\lambda + \tilde{\mu} S(y|a_0)) g'(\lambda + \tilde{\mu} S(y|a_0)).
\end{aligned}
$$

Lemma A.10 states that $\frac{\tilde{\mu}}{\lambda} \to 0$ as $\lambda \to \infty$, and Assumption 4 implies that $(\lambda + \tilde{\mu} S(y_0|a_0)) g'(\lambda + \tilde{\mu} S(y_0|a_0))$ has a finite limit. It follows that

$$
U_{aa}\left(\tilde{V}\left(y|\lambda\right), a\right) \leq -c''(a) < 0
$$

with the last inequality implied by the strict convexity of the cost function (Assumption 1). □

We now prove Proposition 2.

Proof. We first prove that there exists λ_0 such that, for all $\lambda \geq \lambda_0$ and a in $\mathcal{A}$,

$$
U_{aa}(\tilde{V}(y|\lambda), a) \leq 0.
$$

To reach a contradiction, assume that this is not the case. Then there exists a sequence of $\lambda_n \to \infty$ and a_k such that

$$
U_{aa}(\tilde{V}(y|\lambda_n),_n) > 0.
$$

Because $\mathcal{A}$ is compact, we can take a convergent subsequence where $a_k \to a_1$. Therefore,

$$
\limsup_{\lambda \to \infty} U_{aa}(\tilde{V}(y|\lambda), a_1) > 0.
$$

This contradicts Lemma A.8.

For any $\lambda \geq \lambda_0$, we thus have that $U(\tilde{V}(y|\lambda), a)$ is concave in a. The proposition is proven by letting U^* be the solution to $\lambda^*(U^*) = \lambda_0$. The agent's problem is concave in a for $\bar{U} \in [U^*, U_R]$ because $\lambda^*(\bar{U})$ is monotonic by Proposition 1.4 □

We conclude the section by proving two minor lemmas that we used in the above proofs.

Lemma A.9. *There exists y_0 such that for all $y < y_0$,*

$$
\Delta g(y|\lambda) \left(S^2(y|a) + S_a(y|a)\right) \cdot S(Y|a_0) \leq 0.
$$

Proof. The lemma follows from these 3 facts.

1. $\Delta g(y|\lambda) \geq 0$
2. There exists y_0 such that $S(Y|a_0) < 0$ for $y < y_0$
3. There exists y_0 such that $\left(S^2(y|a) + S_a(y|a)\right) > 0$ for $y \leq y_0$.

The monotonicity of g implies the first fact. The second fact follows from Assumption 3.3. Assumption 3.4 and the following algebra implies the final fact.

$$S^2(y|\hat{a}) + S_a(y|a) = S^2(y|\hat{a}) + \frac{d}{da}\frac{f_a(y|a)}{f(y|a)} = S^2(y|\hat{a}) + \frac{f_{aa}(y|a)}{f(y|a)} - \frac{f_a(y|a)^2}{f(y|a)^2} = \frac{f_{aa}(y|a).}{f(y|a)}.$$

The density is always positive, so the final expression is positive whenever $f_{aa}(y|a)$ is positive. ☐

Lemma A.10. *As $\lambda \to \infty$, $\frac{\tilde{\mu}}{\lambda} \to 0$.*

Proof. Proposition A.7 and 3.3 imply that $\underline{S}(\lambda) \to -\infty$. The result follows from the definition of $\underline{S}(\lambda)$. ☐

A.3 Proof of Theorem 1

Parts (1) and (2) follow from Proposition 1. Part (3) follows from Proposition 2. For any $\bar{U} \geq U^*$ we have that $v^*(\bar{U})$ is concave, and thus is a solution to the cost minimization problem. Any other solution also solves the relaxed cost minimization problem, so equals $v^*(\bar{U})$ almost everywhere.