Flexible Moral Hazard Problems*

George Georgiadis[†] *Northwestern University*

Doron Ravid[‡] *University of Chicago*

Balázs Szentes[§] *London School of Economics*

June 1, 2023

Abstract

This paper considers a moral hazard problem where the agent can choose any output distribution with a support in a given compact set. The agent's effort-cost is *smooth* and increasing in first-order stochastic dominance. To analyze this model, we develop a generalized notion of the first-order approach applicable to optimization problems over measures. We demonstrate that each output distribution can be implemented and identify those contracts which implement that distribution. Contracts are characterized by a simple first-order condition which equates the agent's marginal cost of changing the implemented distribution with its marginal benefit. Furthermore, the agent's wage is shown to be increasing in output. Finally, we consider the problem of a profit-maximizing principal and provide a first-order characterization of principal-optimal distributions.

1 Introduction

Perhaps the most celebrated conclusion of the literature on moral hazard is that optimal compensation schemes are designed to reward the agent for those output realizations which are informative about the target level of effort (see, for example, Holmström, 1979

*We would like to thank Henrique Castro-Pires, Daniel Gottlieb, Alessandro Pavan, and Jeroen Swinkels for helpful comments.

[†]Kellogg School of Management, Northwestern University, Evanston, IL 60208, U.S.A., Email: g-georgiadis@kellogg.northwestern.edu, website: https://www.kellogg.northwestern.edu/faculty/georgiadis.

[‡]Kenneth C. Griffin Department of Economics, University of Chicago, Chicago, IL 60637. Email: dravid@uchicago.edu, website: http://www.doronravid.com.

[§]Department of Economics, London School of Economics, London, WC2A 2AE, UK, Email: b.szentes@lse.ac.uk, website: http://personal.lse.ac.uk/szentes.

and 2017). Since larger outputs are not necessarily more informative than smaller ones, optimal wage schemes are often non-monotone in output.[1] These results are typically derived in models in which the action space of the agent is restricted to be either a binary or a one-dimensional set. In this paper, we put forward a model where the agent can *flexibly* choose any output distribution and re-examine the aforementioned conclusions of the literature. We demonstrate that, in such flexible models, optimal wage schemes are not motivated by the informativeness of the output. Instead, they simply compensate the agent for his marginal cost of choosing the target distribution. More precisely, optimal contracts are constructed so that the target distribution satisfies a generalized first-order condition: the agent's marginal cost of choosing a nearby distribution is his marginal benefit from doing so. Moreover, wage schemes are always increasing in output as long as the agent's cost of choosing a distribution is monotone in first-order stochastic dominance.

In the specific model of this paper, there is a single agent. After receiving a wage contract, the agent can choose any output distribution with support in a given compact subset of $\mathbb{R}$. The agent's payoff is additively separable in her utility from wage and the (effort-) cost associated to the selected distribution. Moreover, the agent has limited liability, so the wage must be weakly positive. We make two assumptions on the costs of output distributions. First, the cost is monotone in first-order stochastic dominance. That is, if a distribution first-order stochastically dominates another one then it costs more. Second, this cost is Gateaux differentiable. We explain the notion of Gateaux differentiability in details below. For most of our results, we do not need to specify the principal's preferences. Indeed, our main objective is to derive predictions regarding the wage contracts that incentivise the agent.

In order to illustrate our model and results, it might be useful to consider the following well-known example.

Example 1. Suppose that the agent can choose any distribution with support in $\{0,1\}$. The cost of choosing the distribution which specifies probability p of the output realization one is $c(p)$. The agent's utility from wage is given by the increasing function $u:\mathbb{R}_+\to\mathbb{R}$.

The cost function of this example satisfies our monotonicity and smoothness assumptions whenever c is increasing and differentiable. For each distribution p^*, we next describe

[1] To guarantee that wages are increasing, the distributions available to the agent must satisfy the monotone likelihood property.

those contracts that implement p^*. Fix a wage scheme $w : \{0, 1\} \to \mathbb{R}$ and let m denote the agent's utility from $w(0)$, that is, $m = u(w(0))$. When presented with w, the agent maximizes $pu(w(1)) + (1-p)m - c(p)$ with respect to p. The agent chooses p^* if it satisfies the corresponding first-order condition, that is,

$$u(w(1)) = c'(p^*) + m. \tag{1}$$

For each constant m, the previous equation characterizes a wage scheme that implements p^*. The agent's limited liability constraint determines the smallest m for which such a wage scheme is feasible. There are a number of implications of this observation. First, the principal can implement any distribution p by a wage contract satisfying equation (1). Second, unlike in the classical Holmstrom model, the cost-minimizing wage-scheme is not motivated by the information content of the output. Instead, it simply equates the agent's marginal cost of a distribution with his marginal benefit. Third, the wage scheme is always weakly increasing on the support of the implemented distribution.[2] Our paper demonstrates that all these results generalize to any flexible moral hazard problem as long as the aforementioned two assumptions, monotonicity and smoothness, are satisfied.

Our first main result is that any distribution can be implemented by an appropriate wage schedule. The key to this result is to develop a notion of the first-order approach based on Gateaux differentiability. Roughly speaking, Gateaux differentiability means that the difference between the cost of a given distribution, say μ, and that of another nearby distribution can be well-approximated by the difference between the expectations of a function, c_μ, according to the two distributions. Moreover, the function c_μ depends only on the given distribution μ and it is called the Gateaux derivative of c at μ. We show that a wage scheme, w, implements a distribution μ^*, if the agent's utility from wage is the sum of the Gateaux derivative at μ^* and a constant at each output realization, x, on the support of the distribution and less elsewhere. That is,

$$u(w(x)) = c_{\mu^*}(x) + m$$

for each $x \in \text{supp}(\mu^*)$. Note that this equation generalizes equation (1) of the example. Intuitively, this condition guarantees that the agent has no incentive to modify the target distribution μ^* by relocating probability mass across different output levels. We emphasize that for this condition to hold, the agent's action space does not need to be the entire

[2] If the wage is larger at zero than at one, the agent chooses $p = 0$, so the value one is not in the support of the implemented distribution.

set of distributions. Indeed, it is a necessary condition for implementation as long as the agent can arbitrarily modify the target distribution *locally*.

We consider the main take-away from our analysis to be the observation that, if the agent can choose distributions flexibly, optimal wage contacts are not motivated by the information content of the output. In fact, there is no natural notion of informativeness of the output. To see this, consider a non-generate target distribution μ and an arbitrary output realization x. Then there is a set of distributions under which x is less likely to occur than under μ and there is another set under which x is more likely. In other words, each output realization may indicate the absence of some deviations but suggests the presence of others. Thus, in flexible moral hazard problems, there is no sense in which contracts are designed to reward the agent for those output realizations that are indicative of the target level of effort. Instead, incentive compatible wage schemes must eliminate the agent's desire to relocate probability mass across outputs. To do so, the optimal contract effectively reimburses the agent for the marginal cost of producing each output.

Let us now turn our attention to the monotonicity of the wage schemes. Recall that in standard principal-agent models with hidden action, equilibrium wages are monotone in output only under strong assumptions on the feasible output distributions. In the binary effort case, wages are increasing in output only if the monotone likelihood ratio property is satisfied. If the agent can generate a one-dimensional family of distributions, the wage scheme is increasing only if the derivative of the log-density in effort is increasing. By contrast, in our flexible moral hazard model, the monotonicity of the wage scheme follows directly from the monotonicity of the agent's effort-cost. More precisely, if a wage scheme implements a certain distribution, then this wage scheme is (weakly) increasing on the support of that distribution. This is a rather obvious result and can be explained as follows. Suppose that the wage is larger at a small output level than at other higher outputs. Then the agent would never choose a distribution which specifies positive probability on those higher outputs. The reason is that the agent can modify the distribution by moving the probability mass from those higher outputs to the low output. On the one hand, this modification increases the agent's expected wage because the wage conditional on the low output exceeds the wage conditional on any of those high output levels. On the other hand, the modified distribution is first-order stochastically dominated by the original one, so it is cheaper to the agent.

We conclude our analysis by considering the principal's problem of finding the profit-

maximizing distribution and the corresponding optimal contract. In order to extend the aforementioned first-order approach to the principal's profit-maximization problem, we need to make a stronger smoothness assumption. Roughly speaking, this assumption requires the agent's cost function to be twice differentiable. We then characterize the first-order condition corresponding to the principal's problem. Finally, we illustrate how this first-order condition can be used to derive properties of the principal-optimal distribution. For example, we provide sufficient conditions under which this distribution is degenerate.

Related Literature. First and foremost, our paper is related to the literature on principal-agent problems under moral hazard (Mirrlees, 1976 and Holmström, 1979). In the canonical model the principal offers a wage contract, and then the agent chooses a (typically) one-dimensional action that determines the distribution of output. The optimal contract is shaped by the information content of the output, as well as a trade-off between incentives and insurance. See Holmström (2017) and Georgiadis (2022) for reviews. Instead, the agent can choose *any* output distribution in our model.

Several papers study models where the agent shapes the distribution of the output distribution under semi-parametric assumptions on the cost of distributions.[3] For example, in Diamond (1998) and Barron, Georgiadis, and Swinkels (2020) the agent's cost of choosing a distribution is a function of its mean. Palomino and Prat (2003) assumes that the agent controls the first two moments of the output distribution. In Hébert (2018), who studies security design, costs come from the α-divergence family. This family includes the famous Kullback-Leibler (KL) divergence, which turns out to be the only case in which debt is optimal.[4] Bonham and Riggs-Cragun (2021) and Mattsson and Weibull (2022) also use the KL divergence to model the agent's cost in a moral hazard setting, whereas in Bonham (2021) the cost of assigning probability to each level of output is quadratic.[5,6] In contrast, our model does not impose any functional form assumptions on the cost of each distribution.

[3] A related early contribution is due to Holmstrom and Milgrom (1987) who, in their Section 2, show that with a single performance measure, increasing the dimension of the agent's action space restricts the set of incentive compatible contracts.

[4] Under more general conditions, a mixture of debt and equity is approximately optimal.

[5] Mattsson and Weibull (2022) also derive the agent's first-order condition when costs are given by f-divergence, which includes KL and α-divergence as special cases.

[6] We note that, in addition to their lack of generality, the quadratic and f-divergence costs have the unappealing feature that they do not increase in first-order stochastic dominance.

Our paper is also related to the literature on robust contracting, see for example Carroll (2015), Carroll (2019) for a review, Antic (2022) and Antic and Georgiadis (2022). Like our paper, this literature imposes only minimal restrictions on the technology available to the agent. Their premise, however, is that the principal has limited knowledge regarding the technology and evaluates contracts according to the worst-case scenario. In contrast, the agent's cost of choosing any distribution is common knowledge in our model.

There is also an empirical literature documenting that agents have at least some flexibility of determining distributions over outcomes. For example, Brown, Harlow, and Starks (1996) and Chevalier and Ellison (1997) find that mutual fund managers systematically alter the riskiness of their portfolios in manners consistent with their dynamic incentives. See also Sections 6-7 of Rajan (2011). Shue and Townsend (2017) finds that option grants lead CEOs to take larger risks—most often by increasing leverage, while Rahmandad, Henderson, and Repenning (2018) presents examples where executives sacrifice long-term value to boost short-term profits. Finally, Oyer (1998) and Larkin (2014) present evidence of salespeople manipulating the timing of sales (which is equivalent to manipulating the distribution of sales in each pay-period; see Section 6.3 of Barron, Georgiadis, and Swinkels, 2020).

2 Model

There is an agent who can produce any output distribution with support in a compact subset X of $\mathbb{R}$. Throughout, we let $\underline{x} := \min X$ and $\bar{x} := \max X$ denote the lowest and highest possible outputs, respectively. Let $\mathcal{M}$ denote the set of Borel probability measures on X. The agent's payoff is additively separable in the utility from wage and the effort cost of producing. The utility function from money, $u : \mathbb{R}_+ \to \mathbb{R}$, is strictly increasing, continuous, unbounded and it is normalized so that $u(0) = 0$. The agent's cost of producing $\mu \in \mathcal{M}$ is $C(\mu)$, where $C : \mathcal{M} \to \mathbb{R}_+$ is a weak*-continuous and convex function. So, if the agent chooses $\mu \in \mathcal{M}$ and receives wage w then his payoff is $u(w) - C(\mu)$. Moreover, the agent is an expected-payoff maximizer.

Before the agent decides which distribution to produce, he receives a wage contract. A wage contract is a measurable mapping from realized outputs to monetary compensations. The agent has limited liability so every contract must specify weakly positive wages. To ensure the agent's payoff is well-defined, we also require the agent's contract to be bounded from above. Let $\mathcal{W}$ denote the set of such contracts, that is,

$\mathcal{W} = \{w|\ w : X \to \mathbb{R}_+, \sup w(X) < \infty\}$.

We next argue that assuming the convexity of C is without loss. Indeed, since the agent may randomize, the cost of any distribution should be evaluated by the expected cost of the cheapest randomization that generates it, resulting in a convex cost function. We state two further assumptions on the cost of production. First, we assume that producing more in the sense of first-order stochastic dominance costs more.

Assumption 1. (**monotonicity**) If the distribution μ first-order stochastically dominates μ' then $C(\mu) \geq C(\mu')$.

Our second assumption ensures that the cost function is smooth.

Assumption 2. (smoothness) The function C is **Gateaux differentiable**, which means that every μ admits a continuous function $c_\mu : [0,1] \to \mathbb{R}$ such that

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \left[C(\mu + \epsilon(\mu' - \mu)) - C(\mu) \right] = \int c_\mu(x) \, (\mu' - \mu) \, (\mathrm{d}x)$$

for all $\mu' \in \mathcal{M}$. The function c_μ is referred to as the (Gateaux) derivative of C. [7]

Let us make a few remarks regarding Assumption 2. First, if c_μ is a derivative of C at μ, then so is $c_\mu + k$ for any constant $k \in \mathbb{R}$. It is therefore without loss to require $c_\mu(\underline{x}) = 0$. Second, whenever Assumption 2 holds, Assumption 1 is equivalent to c_μ being increasing for all μ (see Cerreia-Vioglio, Maccheroni, and Marinacci, 2017, for example). And third, when there are only n outputs, $X = \{x_1, \ldots, x_n\}$, C becomes a mapping from the n-dimensional simplex to $\mathbb{R}_+$. In this case, Assumption 2 is equivalent to the usual notion of differentiability, and one can express c_μ in terms of the partial derivatives of C. More specifically, let $C_i'(\mu)$ denote the partial derivative of C with respect to the probability of output i at the distribution μ, and suppose $x_1 = \underline{x}$ is the lowest output. Then one can express the Gateaux derivative of C as $c_\mu(x_i) = C_i'(\mu) - C_1'(\mu)$.

Our goal is to analyze the set of those distributions which can be implemented and characterize the wage contracts which implement them. More formally, for each $w \in \mathcal{W}$, the measure $\mu \in \mathcal{M}$ is called w**-incentive compatible** (w-IC) if the agent finds it optimal to produce μ after he receives the contract w. Note that if the wage contract is w and

[7] We assume c_μ is continuous to be consistent with the definition of Gateaux differentiability in the related decision-theoretic literature (e.g., Hong, Karni, and Safra, 1987; Cerreia-Vioglio, Maccheroni, and Marinacci, 2017). Our results continue to hold if we relaxed Assumption 2 to require c_μ only to be lower semicontinuous and have finite integral under μ.

the agent chooses $\mu \in \mathcal{M}$, then his payoff is

$$U(\mu, w) = \int u \circ w(x) \mu \, (\mathrm{d}x) - C(\mu).$$

So, the measure μ is w-IC if $U(\mu, w) = \sup_{\mu' \in \mathcal{M}} U(\mu', w)$. We say μ is **implementable** whenever it is w-IC for some $w \in \mathcal{W}$.

We emphasize that for most of our results, we do not need both assumptions above. For example, even if neither of these assumptions hold, the set of implementable distributions is large.

Theorem 1 *The set of distributions that is implementable is dense.*

Proof. See the Appendix. ■

To prove the above theorem, we identify each measure in $\mathcal{M}$ with its corresponding CDF. By equipping the set of CDFs with the L^2-norm, we recast C as a convex and lower-semicontinuous function over a Banach space. To conclude the proof, we show one can implement every CDF at which the subdifferential of C is non-empty, a condition which holds over a dense set of the cost function's domain by the Brondsted-Rockafellar theorem (Brøndsted and Rockafellar, 1965).

We conclude this section by providing an example for the agent's cost function, C, which satisfies our assumptions. We will use this example to illustrate many of our results throughout the paper.

Example 2. Let $X \subseteq \mathbb{R}$ be any compact set and $c : X \to \mathbb{R}_+$ an increasing and continuous function with $c(\underline{x}) = 0$. Furthermore, let $K : \mathbb{R} \to \mathbb{R}$. be an increasing, convex, and differentiable function. If the agent's cost function is defined by

$$C(\mu) = K\left(\int c(x) \ \mu(\mathrm{d}x)\right), \tag{2}$$

it satisfies Assumptions 1 and 2. Indeed, by the Chain Rule, this function is Gateaux differentiable, with the derivative given by

$$c_\mu(x) = K'\left(\int c(y) \ \mu(\mathrm{d}y)\right) c(x). \tag{3}$$

3 Main Results

3.1 Monotone Wages

Our first result establishes that the monotonicity of C (Assumption 1) implies the monotonicity of any wage scheme on the support of the distribution it implements.

Definition 1 *The contract $w \in \mathcal{W}$ is μ-**almost increasing** if for all $x \in X$,*

$$\mu\left(\left\{x' \in X : x < x', w\left(x'\right) < w\left(x\right)\right\}\right) = 0.$$

We are ready to state our first result.

Proposition 1 *If Assumption 1 holds and $\mu \in \mathcal{M}$ is w-IC then w is μ-almost increasing.*

The proof of the proposition is established along the same arguments described in the Introduction. We show that if w is not μ-almost increasing, then the measure μ can be modified by moving probability from high outputs at which the wage is low to a low output realization at which the wage is high. This new measure then is cheaper to the agent and generates higher expected utility, that is, μ was not w-IC.

Proof. Suppose, by contradiction, that w is not μ-almost increasing. Then there exists $x \in X$ such that $\mu\left(S_x\right) > 0$, where

$$S_x = \left\{x' \in X : x' > x, w\left(x'\right) < w\left(x\right)\right\}.$$

Let $\mu' \in M$ be a modification of μ so that all the mass from the set S_x is moved to x. Formally, for each Borel set A,

$$\mu'\left(A\right) = \begin{cases} \mu\left(A \backslash S_x\right) + \mu\left(S_x\right) & \text{if } x \in A, \\ \mu\left(A \backslash S_x\right) & \text{otherwise.} \end{cases}$$

Since $x < x'$ for all $x' \in S_x$ and $\mu\left(S_x\right) > 0$, it follows that μ strictly first-order stochastically dominates μ'. Finally, note that

$$U(\mu, w) \leq \int u \circ w(x) \mu\left(\mathrm{d}x\right) - C(\mu') < \int u \circ w(x) \mu'\left(\mathrm{d}x\right) - C(\mu') = U(\mu', w),$$

where the weak inequality follows from Assumption 1 and the fact that μ first-order stochastically dominates μ' and the strict inequality follows because $w\left(x\right) > w\left(x'\right)$ for each $x' \in S_x$ and $\mu\left(S_x\right) > 0$. This inequality chain implies that $U(\mu, w) < \sup_{\mu' \in \mathcal{M}} U(\mu', w)$, that is, μ is not w-IC, a contradiction. ■

3.2 Implementability

Next, we explore the consequences of Assumption 2. The following lemma develops a notion of the first-order approach based on Gateaux differentiability. In particular, it proves necessity and sufficiency of a first-order condition for maximization. The first-order

approach is then applied to characterize the agent's optimal distribution for a given wage contract. In turn, this leads to our main result: each distribution can be implemented[8] and the corresponding wage scheme is determined by the aforementioned first-order condition.

To understand how the statement of the next lemma is related to the first-order condition familiar from one-dimensional calculus, consider the problem of maximizing $vx - c(x)$ on $[0,1]$, where $v \in \mathbb{R}_+$, and c is a convex, differentiable function. Then $x^* \in (0,1)$ solves this problem if and only if it satisfies the first-order condition $v = c'(x^*)$. An equivalent way of stating it is that $x^* \in (0,1)$ solves the problem if and only if x^* also solves $\max_{x\in[0,1]}(vx - c'(x^*)x)$. In what follows, we generalize this latter condition for Gateaux differentiable cost functions.

Lemma 1 *For a bounded and measureable $v : X \to \mathbb{R}$, and $\mu^* \in \mathcal{M}$,*

$$\mu^* \in \operatorname*{argmax}_{\mu\in\mathcal{M}} \int v(x)\mu\,(\mathrm{d}x) - C(\mu)$$

if, and only if

$$\mu^* \in \operatorname*{argmax}_{\mu\in\mathcal{M}} \int \left(v(x) - c_{\mu^*}(x)\right)\mu\,(\mathrm{d}x)\,.$$

We note that the convexity of the function C plays a role only in the "if" part of the proof. That is, the first-order condition would be necessary even if C was not convex.[9]

Proof. We first prove that the first order-condition is necessary. Fix any $\tilde{\mu} \in \mathcal{M}$. For all $\epsilon \in (0,1)$, define $\mu_\epsilon := \mu^* + \epsilon(\tilde{\mu} - \mu^*)$, which is in the convex set $\mathcal{M}$. If $\mu^* \in \operatorname{argmax}_{\mu\in\mathcal{M}}\left[\int v(x)\mu\,(\mathrm{d}x) - C(\mu)\right]$ then

$$0 \geq \frac{1}{\epsilon}\left[\int v(x)\ (\mu_\epsilon - \mu^*)(\mathrm{d}x)\right] - \frac{1}{\epsilon}\left[C(\mu_\epsilon) - C(\mu^*)\right] = \int v(x)\ (\tilde{\mu} - \mu^*)(\mathrm{d}x) - \frac{1}{\epsilon}\left[C(\mu_\epsilon) - C(\mu^*)\right]$$

where the inequality follows from μ^* being a maximizer and the equality is implied by the definition of μ_ϵ. Observe that, since C is Gateaux differentiable at μ^*, the last expression of the previous displayed inequality chain converges to

$$\int \left[v(x) - c_{\mu^*}(x)\right]\ (\tilde{\mu} - \mu^*)(\mathrm{d}x),$$

as ϵ goes to zero.

[8] Recall that Theorem 1 only states that, absent Assumption 2, the set of implementable distributions is dense.

[9] We also note that an identical proof shows the lemma continues to hold if one replaces $\mathcal{M}$ with any convex subset, $\bar{\mathcal{M}} \subseteq \mathcal{M}$.

We now show that the first-order conditionis sufficient when C is convex. To that end, we first claim that

$$C(\mu) - C(\mu^*) \geq \int c_{\mu^*}(x)\ (\mu - \mu^*)(\mathrm{d}x) \tag{4}$$

holds for all μ. To prove this inequality, note that the convexity of C means that

$$\frac{1}{\epsilon}\left[C(\mu^* + \epsilon(\mu - \mu^*)) - C(\mu^*)\right]$$

is decreasing in $\epsilon \in (0,1)$. Letting $(\epsilon_n)_{n\in\mathbb{N}}$ be a decreasing sequence in $(0,1)$ converging to zero, we have

$$C(\mu) - C(\mu^*) \geq \frac{1}{\epsilon_n}\left[C(\mu^* + \epsilon_n(\mu - \mu^*)) - C(\mu^*)\right] \xrightarrow{n\to\infty} \int c_{\mu^*}\ \mathrm{d}(\mu - \mu^*).$$

Therefore, if μ^* satisfies the first order condition, the following must hold for every μ:

$$0 \geq \int \left(v - c_{\mu^*}\right)(x)\left(\mu - \mu^*\right)(\mathrm{d}x) \geq \int v(x)\left(\mu - \mu^*\right)(\mathrm{d}x) - [C(\mu) - C(\mu^*)],$$

where the first inequality follows from the fact that μ^* satisfies the first-order condition, that is, $\mu^* \in \operatorname{argmax}_{\mu\in\mathcal{M}}\left[\int (v(x) - c_{\mu^*}(x))\,\mu\,(\mathrm{d}x)\right]$. The second inequality is just (4). Finally, the previous inequality chain implies $\mu^* \in \operatorname{argmax}_{\mu\in\mathcal{M}}\left[\int v(x)\mu\,(\mathrm{d}x) - C(\mu)\right]$. ■

Next, we apply the previous lemma to the agent's problem of choosing a distribution. To this end, for each $\mu \in \mathcal{M}$, let $m^*(\mu) = \inf\{m : \min_{x\in X} c_\mu(x) + m \geq 0\}$ and for each $m \geq m^*(\mu)$, define

$$w_{\mu,m}(x) := u^{-1}\left(c_\mu(x) + m\right).$$

The next proposition states that the wage contract $w_{\mu,m}$ implements μ for each $m \geq m^*$.

Proposition 2 *Suppose that C satisfies Assumption 2. Then, the measure $\mu \in \mathcal{M}$ is w-IC if, and only if,*

$$w(x)\begin{cases} = w_{\mu,m}(x) & \text{if } x \in Y, \\ \leq w_{\mu,m}(x) & \text{otherwise,} \end{cases}$$

holds for some $m \geq m^*(\mu)$ and some $Y \subseteq X$ with $\mu(Y) = 1$.

Proof. Observe that the agent's objective, $\int u \circ w(x)\mu\,(\mathrm{d}x) - C(\mu')$, is concave and Gateaux differentiable in μ'. Therefore, Lemma 1 implies that μ is w-IC if and only if μ satisfies the agent's first-order condition, that is, μ solves

$$\max_{\mu'} \int \left(u \circ w(x) - c_\mu(x)\right)\mu'\,(\mathrm{d}x).$$

which is equivalent to $u \circ w(x) - c_\mu(x) \leq \sup_{x \in [0,1]} u \circ w(x) - c_\mu(x) =: m$ holding with equality μ-almost surely. The proposition follows from rearranging this inequality and noting that u^{-1} is strictly increasing. Finally, note that $m \geq m^*(\mu)$ must hold because of limited liability. ■

We now show that the previous proposition implies that every $\mu \in \mathcal{M}$ can be implemented, and that the contract

$$w^*_\mu := w_{\mu, m^*(\mu)}$$

is a cost-minimizing contract among those that implement μ.

Corollary 1 *Suppose C satisfies Assumption 2, and fix any $\mu \in \mathcal{M}$. Then μ is w^*_μ-IC. Moreover, for any other $w \in \mathcal{W}$ for which μ is w-IC,*

$$w \geq w^*_\mu$$

holds μ-almost surely.

We point out that the cost-minimizing wage scheme implementing any μ is uniquely determined μ-almost everywhere. For sets that arise with zero probability under μ, the cost-minimizing contract can be defined arbitrarily as long as it is weakly smaller than w^*_μ.

Proof. That μ is w^*_μ-IC follows immediately from Proposition 2. The same proposition also implies that every $w \in \mathcal{W}$ for which μ is w-IC, there exists some $m \geq m^*(\mu)$ such that $w = w_{\mu,m}$ μ-almost surely. Since $w_{\mu,m}(x) = u^{-1}(c_\mu(x) + m)$, $m \geq m^*(\mu)$, and u^{-1} is strictly increasing, it follows that $w_{\mu,m} \geq w_{\mu,m^*(\mu)} = w^*_\mu$. ■

To conclude this section, we consider what happens when the cost function C satisfies both Assumptions 1 and 2. We show that, in this case, one can obtain a more complete characterization of a cost-minimizing contract and that it can be chosen to be monotone everywhere.

Corollary 2 *Suppose C satisfies Assumptions 1 and 2. Then,*

$$w^*_\mu(x) = u^{-1}(c_\mu(x)).$$

is a cost-minimizing contract among those that implement μ. Moreover, w^*_μ is increasing everywhere.

Recall that Proposition 1 states that any wage schedule is increasing almost everywhere on the support of the measure it implements. This corollary says that if Assumption 2 also holds then, for each measure μ, there is a cost-minimizing wage which implements μ and is increasing everywhere, even outside the support of μ.

Proof. We first note that Assumption 1 implies that c_μ is increasing, see (e.g., Cerreia-Vioglio, Maccheroni, and Marinacci, 2017).[10] Since u^{-1} is also increasing, $u^{-1}(c_\mu(x) + m) \geq 0$ if and only this inequality holds at $x = 0$. Therefore, the normalizations, $u(0) = 0$ and $c_\mu(\underline{x}) = 0$, imply that $m^*_\mu = 0$. Consequently, $w^*_\mu(x) = u^{-1}(c_\mu(x))$ and this function is increasing. ■

We now revisit the example of Section 2 and compute the cost-minimizing wage contract for each distribution. We also show that in the special case where the agent's marginal cost is constant, the cost-minimizing wage does not depend on the implemented distribution.

Example 2. (continued.) Recall that the agent's cost function is defined by (2) and its Gateaux derivative is given by (3). Moreover, this cost function is monotone. Therefore, an immediate consequence of the previous corollary is that, for each μ, the cost minimizing wage is given by the following equation:

$$w^*_\mu(x) = u^{-1}\left(K'\left(\int c(y)\ \mu(\mathrm{d}y)\right) c(x)\right). \tag{5}$$

This equation has some interesting implications in the special case where the agent's marginal cost is constant, that is, the function K is the identity function. Since $K' = 1$, equation (5) simplifies to $w^*_\mu(x) = u^{-1}(c(x))$ for *all* output distributions. In other words, this wage scheme is the cost-minimizing for each distribution. Notice this contract results in the agent getting a net utility of zero regardless of the output, since

$$u(w^*_\mu(x)) - c(x) = u(u^{-1}(c(x))) - c(x) = 0.$$

4 Profit Maximization

In this section, we turn our attention to the principal's problem of finding the profit-maximizing distribution and the corresponding contract. We assume the principal's payoff

[10] If one wishes to relax Assumption 2 in a way that allows c_μ to be lower-semicontinuous (see footnote 7), one cannot apply the referenced result. We therefore provide an independent proof of this fact in the appendix.

is $x - w$ if the output is x and she pays wage w to the agent, and that she is an expected payoff-maximizer. We first make a further assumption on the cost function C which roughly requires it to be twice differentiable. Then we show that a consequence of this assumption is that the principal's profit as a function of the implemented distribution μ is also Gateaux differentiable and characterize a first-order condition corresponding to the principal's problem. Finally, we illustrate how this first-order condition can be used to make meaningful statements about the principal-optimal distribution and contract.

Let us now state the aforementioned assumption which essentially requires the Gateaux derivative of C to be Gateaux differentiable.

Assumption 3. The cost function is Gateaux differentiable, with $\mu \mapsto c_\mu(\cdot)$ being weak*-to-supnorm continuous. Moreover, for every μ, a continuous function $h_\mu : X \times X \to \mathbb{R}$ exists such that for all $\tilde{\mu} \in \mathcal{M}$,

$$\frac{1}{\epsilon}\left[c_{\mu+\epsilon(\tilde{\mu}-\mu)}(\cdot) - c_\mu(\cdot)\right] \xrightarrow{\epsilon \searrow 0} \int h_\mu(\cdot, y)(\tilde{\mu} - \mu)(\mathrm{d}y),$$

where convergence is according to the sup norm, $\|\cdot\|_\infty$.

Let us now describe the problem of a profit-maximizing principal. In order to maximize her profit, the principal chooses an output distribution and a wage contract which implements it. Formally, the principal's program can be written as

$$\max_{\mu \in \mathcal{M}, w \in \mathcal{W}} \int [x - w(x)]\mu(\mathrm{d}x), \text{ subject to } \mu \text{ is } w\text{-IC}.$$

Of course, if a pair (μ, w) solves this problem then the wage scheme w is cost-minimizing among those that implement μ. For each $\mu \in \mathcal{M}$, let $W(\mu)$ be the expected cost-minimizing wage implementing μ.[11] Then, the principal's program can be rewritten as

$$\max_{\mu \in \mathcal{M}} \left[\int x \ \mu(\mathrm{d}x) - W(\mu)\right]. \tag{6}$$

We call a distribution μ^* *principal-optimal*, if it solves this maximization problem.

We aim to provide a partial characterization of a principal-optimal distribution in two steps. First, we compute the Gateaux derivative of the function W. And second, we appeal to Lemma 1 to derive a necessary first-order condition for μ to be principal-optimal.

[11] Recall that, by Corollary 2, $W(\mu) = \int [w^*_\mu(x)]\mu(\mathrm{d}x)$.

To this end, suppose u' is strictly positive, and define the function $\kappa_\mu^* : X \to \mathbb{R}$ as follows

$$\kappa_\mu^*(x) = \int \left(\frac{h_\mu(y,x)}{u'\left(w_\mu^*(y)\right)} \right) \mu(\mathrm{d}y).$$

To interpret κ_μ^*, note that $h_\mu(y,x)$ represents the change in the marginal cost of producing output y associated with a slight increase in the probability of output x. Multiplying $h_\mu(y,x)$ by the ratio $1/u'(w_\mu^*(y))$ converts the change in the agent's marginal cost to a change in the agent's monetary wage. Thus, $\kappa_\mu(x)$ gives the marginal change in the agent's expected compensation associated with an increase in the probability of output x.

The next theorem describes the Gateaux derivative of the principal's expected-wage payments under the cost-minimizing contract as a function of the induced output distribution.

Lemma 2 *Suppose C satisfies Assumptions 1 and 3 and u is a continuously differentiable function with a strictly positive derivative. Then the function W is continuous and Gateaux differentiable with derivative*

$$w_\mu^*(x) + \kappa_\mu^*(x).$$

Each term in the Gateaux derivative, $w_\mu^*(x) + \kappa_\mu^*(x)$, expresses a different force that impacts the principal's expected payments when she shifts the implemented output distribution away from μ. The first term, $w_\mu^*(x)$, is the wage the agent receives when generating an output of x. The second term, $\kappa_\mu^*(x)$, expresses how the change in the cost-minimizing contract that arises due to a change in the probability of x impacts the agent's expected compensation under μ.

We are now ready to characterize the first-order condition describing a principal-optimal distribution. Recall that Lemma 1 developed a first-order approach for a class of maximization problems. Substituting $v(x) = x$ and $C(\mu) = W(\mu)$ into the statement of the lemma, and noting that W is Gateaux differentiable (by Lemma 2), it becomes clear that the "if" part of the statement of Lemma 1 is applicable to the principal's profit maximization problem (6). Since W may not be convex, the "only if" part is not applicable, so the following theorem provides only a necessary condition for optimality.

Theorem 2 *Suppose C satisfies Assumptions 1 and 3, and that u is a continuously differentiable function with a strictly positive derivative. Then, a principal-optimal μ^* exists and*

$$\operatorname{supp} \mu^* \subseteq \operatorname*{argmax}_{x \in X} \left[\pi_{\mu^*}(x) \right], \tag{7}$$

where $\pi_\mu(x) := x - w_\mu^*(x) - \kappa_\mu^*(x)$.

Proof. Note that, by Lemma 2, the principal's objective function in (6) is continuous. Since the domain $\mathcal{M}$ is compact, the existence of a principal-optimal distribution follows. As mentioned above, equation (7) is implied by the "if" part of the statement by Lemma 1. ■

Let us return to Example 2 to illustrate how to compute the derivative of the principal's expected profit for each distribution.

Example 2.(continued.) Recall that C is given by (2) and that we have already characterized w_μ^*, see equation (5). By the previous theorem, in order to derive the derivative of the principal's expected profit, it remains to compute κ_μ^*. To this end, assume that K is twice continuously differentiable. Then C also satisfies Assumption 3, where

$$h_\mu(x,y) = K''\left(\int c(z)\ \mu(\mathrm{d}z) \right) c(x)c(y).$$

Furthermore, whenever u is a continuously differentiable function with a strictly positive derivative,

$$\kappa_\mu^*(x) = K''\left(\int c(z)\mu(\mathrm{d}z) \right) \left[\int \frac{c(y)}{u'\left(w_\mu^*(y)\right)} \mu(\mathrm{d}y) \right] c(x). \tag{8}$$

Let us now return to the general analysis and demonstrate that the condition in (7) can be used to deduce properties of the principal's optimal distribution and the corresponding wage contract. Observe that this condition depends on the function π_μ, which we characterized in terms of the agent's utility function u and cost function C. The next corollary establishes relationships between the shape of π_μ and the support of the optimal distribution.

Corollary 3 *Suppose Assumptions 1 and 3 hold, $X = [\underline{x}, \overline{x}]$ and that u is a continuously differentiable function with a strictly positive derivative.*

(i) If π_μ is strictly quasiconcave for every μ with more than one output, the principal optimal distribution has at most one output in its support.

(ii) If π_μ is strictly quasiconvex for every μ that includes some non-extreme output $x \in (\underline{x}, \bar{x})$ in its support, the principal optimal distribution is supported on $\{\underline{x}, \bar{x}\}$.

(iii) If $w_\mu + \kappa_\mu$ is a non-affine analytic function whenever μ is not discrete, the principal optimal distribution is discrete.

Proof. As explained above, if μ is principal optimal, it must be supported on the set of outputs that maximize π_μ. Part (i) then follows from observing that this set can have at most one output whenever π_μ is strictly quasiconcave. Part (ii) follows from noting that a strictly quasiconvex function over a compact interval is maximized at the interval's end points. For Part (iii), observe that $w_\mu + \kappa_\mu$ being a non-affine analytic function means the function $x \mapsto [\pi_\mu(x) - \max \pi_\mu(X)]$ is a non-zero analytical function. Therefore, by the identity theorem, the set

$$\operatorname*{argmax}_{x \in X}[\pi_\mu(x)] = \{x \in X : \pi_\mu - \max \pi_\mu(X) = 0\}$$

cannot have any accumulation points in $(\underline{x}, \bar{x})$. The conclusion follows. ■

Let us illustrate each part of the previous corollary by considering various specifications of Example 2.

Example 2.(continued.) Note that Theorem 2 implies the derivative of the principal's expected profit is $\pi_\mu(x) = x - w^*_\mu(x) - \kappa^*_\mu(x)$, where w^*_μ and κ^*_μ are given by (5) and (8), respectively. Suppose X is an interval, the agent is risk neutral, and $u(x) = x$, so $u'(\cdot) = 1$. Then if c is strictly convex, π_μ is strictly concave (hence strictly quasiconcave), and so part (i) implies it is always optimal to induce a single output. If c is strictly concave instead, part (ii) implies the principal optimal distribution has at most two outputs, because π_μ is strictly (quasi-)convex. Finally, if we replace the convexity or concavity assumptions with the postulate that c is a non-affine analytic function, the same holds for π_μ, in which case the principal optimal distribution must be discrete by part (iii).

Corollary 3 is particularly useful when either part (i) or part (ii) hold. In these cases, the principal's program reduces to a one dimensional optimization problem. To see this, suppose first that Corollary 3's assumptions hold and that π_μ is strictly quasiconcave for every μ. By Corollary 3, the principal optimal distribution has only one output. Letting δ_x be the distribution generating output x with probability 1, it follows the principal optimal output distribution must solve

$$\max_{x \in X} [x - w_{\delta_x}(x)]. \tag{9}$$

Suppose now instead that π_μ is strictly quasiconvex for every μ. Applying Corollary 4, we obtain the principal optimal distribution takes the form $\mu_p := p\delta_{\bar{x}} + (1-p)\delta_{\underline{x}}$ for some $p \in [0,1]$, and so one can write the principal's problem as

$$\max_{p\in[0,1]} \left[(1-p)\underline{x} + p\bar{x} - pw_{\mu_p}(\bar{x})\right].^{12} \tag{10}$$

Hence, the principal's problem reduces to finding the optimal probability p with which to generate the highest output, just as in the binary output example.

Finally, we reconsider our running example with a risk-averse agent.

Example 2. (continued.) Suppose $\underline{x} = 0$, $c(x) = x^\gamma$, $K(a) = a^{1+\lambda}/(1+\lambda)$, and $u(y) = y^\rho$ for γ, λ, and ρ all strictly positive, and $\rho < 1$. In this case, simple algebra reveals that $w^*_\mu(x)$ equals a positive constant times $x^{\gamma/\rho}$, whereas $\kappa^*_\mu(x)$ is some positive constant times x^γ, with both constants being strictly positive whenever $\mu \neq \delta_0$. Hence, if $\gamma \leq \rho$, $\pi_\mu(x) = x - (w^*_\mu(x) + \kappa^*_\mu(x))$ is strictly convex for all $\mu \neq \delta_0$, and so (by Corollary 3, part (ii)) the optimal distribution takes the form $\mu_p = p\delta_{\bar{x}} + (1-p)\delta_0$, where p solves the program detailed in (10). If $\bar{x} = 1$, then $w_{\mu_p}(0) = 0$ and $w_{\mu_p}(1) = p^{\lambda/\rho}$, and so the principal's program becomes

$$\max_{p\in[0,1]} \left[p - p^{\frac{\lambda+\rho}{\rho}}\right].$$

Clearly, the above objective is concave, and so one can solve for the optimal p using the principal's first order condition, the solution to which is

$$p^* = \left[\frac{\rho}{\lambda+\rho}\right]^{\frac{\rho}{\lambda}}.$$

If $\gamma \geq 1 > \rho$, π_μ is strictly concave, and so part (i) of Corollary 3 implies the optimal distribution induces a single output x^*, which is determined by the program in (9). The objective in this program is given by $x - x^{\frac{\gamma}{\rho}(1+\lambda)}$. Since $\gamma > \rho$ and $\lambda > 0$, this objective is strictly concave, and so the optimal x^* is equal to the lower of $\bar{x}$ and the solution to the first order condition; that is, $x^* = \min\{\bar{x}, [\rho/\gamma(1+\lambda)]^{\rho/[\gamma(1+\lambda)-\rho]}\}$.

Finally, if $\rho < \gamma < 1$, π_μ is neither always concave nor always convex. However, it is apparent that the function $w^*_\mu + \kappa^*_\mu$ is a non-affine analytical function whenever μ assigns positive probability to any output strictly larger than 0. As such, one can apply part (iii) of Corollary 3 to obtain that, in this case, the principal optimal distribution must be discrete.

[12] Recall $w_\mu(\underline{x}) = u^{-1}(c_\mu(\underline{x})) = u^{-1}(0) = 0$.

5 Conclusion

Our goal in this paper was to explore the consequences of the agent's flexibility in generating output distributions in moral hazard problems. Our analysis is based on a generalized notion of the first-order approach. We demonstrated that, unlike in the classical model, the cost-minimizing contract is not motivated by the information content of the output regarding the target distribution. Instead, optimal contracts are constructed so that the target distribution satisfies a simple first-order condition which equates the agent's marginal cost of changing the distribution locally with its marginal benefit. We also showed that optimal wage contracts are monotone whenever the agent's cost function is increasing in first-order stochastic dominance. Finally, we applied our first-order approach to the principal's profit maximization problem and provided a partial characterization of principal-optimal output distributions.

References

Antic, Nemanja. 2022. "Contracting with unknown technologies." .

Antic, Nemanja and George Georgiadis. 2022. "Robust Contracts: A Revealed Preference Approach." .

Barron, Daniel, George Georgiadis, and Jeroen Swinkels. 2020. "Optimal Contracts with a Risk-taking Agent." *Theoretical Economics* 15 (2):715–761.

Bonham, Jonathan. 2021. "Shaping Incentives through Measurement and Contracts." .

Bonham, Jonathan and Amoray Riggs-Cragun. 2021. "Contracting on what Firm Owners Value." .

Brøndsted, Arne and Ralph Tyrrell Rockafellar. 1965. "On the Subdifferentiability of Convex Functions." *Proceedings of the American Mathematical Society* 16 (4):605–611.

Brown, Keith C, W Van Harlow, and Laura T Starks. 1996. "Of Tournaments and Temptations: An Analysis of Managerial Incentives in the Mutual Fund Industry." *The Journal of Finance* 51 (1):85–110.

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

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

Cerreia-Vioglio, Simone, Fabio Maccheroni, and Massimo Marinacci. 2017. "Stochastic Dominance Analysis Without the Independence Axiom." *Management Science* 63 (4):1097–1109.

Chevalier, Judith and Glenn Ellison. 1997. "Risk taking by Mutual Funds as a Response to Incentives." *Journal of Political Economy* 105 (6):1167–1200.

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

Georgiadis, George. 2022. "Contracting with Moral Hazard: A Review of Theory & Empirics." .

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

Holmström, Bengt. 1979. "Moral Hazard and Observability." *The Bell Journal of Economics* :74–91.

———. 2017. "Pay for Performance and Beyond." *American Economic Review* 107 (7):1753–77.

Holmstrom, Bengt and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* :303–328.

Hong, Chew Soo, Edi Karni, and Zvi Safra. 1987. "Risk Aversion in the Theory of Expected Utility with Rank Dependent Probabilities." *Journal of Economic theory* 42 (2):370–381.

Larkin, Ian. 2014. "The cost of high-powered incentives: Employee gaming in enterprise software sales." *Journal of Labor Economics* 32 (2):199–227.

Mattsson, Lars-Göran and Jorgen Weibull. 2022. "An Analytically Solvable Principal-Agent Model." *Available at SSRN 4252495* .

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

Oyer, Paul. 1998. "Fiscal year ends and nonlinear incentive contracts: The effect on business seasonality." *The Quarterly Journal of Economics* 113 (1):149–185.

Palomino, Frédéric and Andrea Prat. 2003. "Risk Taking and Optimal Contracts for Money Managers." *RAND Journal of Economics* :113–137.

Rahmandad, Hazhir, Rebecca Henderson, and Nelson P Repenning. 2018. "Making the Numbers? Short Termism and the Puzzle of only Occasional Disasters." *Management Science* 64 (3):1328–1347.

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

Shue, Kelly and Richard R Townsend. 2017. "How do Quasi-random Option Grants affect CEO Risk-taking?" *The Journal of Finance* 72 (6):2551–2588.

Wang, Tan. 1993. "Lp-Fréchet differentiable preference and" local utility" analysis." *Journal of Economic Theory* 61 (1):139–159.

A Proofs Appendix

Proof of Theorem 1.

We begin with some notation. Let $\bar{\mathcal{F}}$ be the set of CDFs over $\bar{X} = \text{co}X = [\underline{x}, \bar{x}]$, endowed with the topology of convergence in distribution, and $\bar{\mathcal{M}}$ the set of Borel measures over $\bar{X}$, endowed with its weak* topology. It is well known that the mapping taking every $F \in \mathcal{F}$ to its induced measure μ_F—i.e., the measure such that $\mu_F[0,x] = F(x)$ for all x—is a linear homeomorphism between $\bar{\mathcal{F}}$ and $\bar{\mathcal{M}}$. By Theorem 1 of Wang (1993), $\bar{\mathcal{F}}$ can be viewed as a subspace of the Banach space $L^2\left(\bar{X}, \lambda\right)$, where λ is the Lebesgue measure. Let $\mathcal{F}$ be the set of CDFs whose support is contained in X, and define the function

$$\hat{C} : L^2\left(\bar{X}, \lambda\right) \to \mathbb{R} \cup \{\infty\}$$
$$\phi \mapsto \begin{cases} C\left(\mu_\phi\right) & \text{if } \phi \in \mathcal{F} \\ \infty & \text{otherwise.} \end{cases}$$

Given a CDF $F \in \mathcal{F}$, define the subdifferential of $\hat{C}$ at F as

$$\partial\hat{C}(F) = \left\{\phi \in L^2\left(\bar{X}, \lambda\right) : \hat{C}(\varphi) \geq \hat{C}(F) + \int \phi(x)(\varphi - F)(x)\,\lambda(\mathrm{d}x)\ \forall \varphi \in L^2\left(\bar{X}, \lambda\right)\right\}.$$

In general, $\partial\hat{C}$ might be empty. Let $\mathcal{F}_I = \left\{F \in \bar{\mathcal{F}} : \partial\hat{C}(F) \neq \varnothing\right\}$ be the set of all CDFs at which $\partial\hat{C}$ is non-empty. Since $\mathcal{F}$ is convex, and C is convex and continuous, it follows $\hat{C}$ is convex and lower semicontinuous. Noting $\hat{C}$ is also proper, it follows from the Brondsted-Rockafellar Theorem (Brøndsted and Rockafellar, 1965) that $\mathcal{F}_I$ is dense in $\mathcal{F}$.

Given $\mu \in \mathcal{M}$, define F_μ to be the CDF such that $\mu_{F_\mu} = \mu$. To conclude the proof, we argue μ is implementable whenever $F_\mu \in \mathcal{F}_I$ (observe this set is dense due to $\mathcal{F}$ and $\mathcal{M}$ being homeomorphic). Indeed, let $\phi \in \partial\hat{C}(F_\mu)$, and define $\Phi(x) := \int_0^x \phi(\tilde{x})\,\mathrm{d}\tilde{x}$, where the integral is viewed as a Riemann-Stieltjes integral. Note

$$w(x) := u^{-1}\left(\max \Phi(X) - \Phi(x)\right)$$

is well-defined because $\max \Phi(X) - \Phi(x) \in u(\mathbb{R}_+)$ for all x. Then for every $\mu' \in \mathcal{M}$,

$$\begin{aligned} C\left(\mu'\right) = \hat{C}\left(F_{\mu'}\right) &\geq \hat{C}(F_\mu) + \int \phi(x)\left(F_{\mu'} - F_\mu\right)(x)\,\mathrm{d}x \\ &= \hat{C}(F_\mu) - \int \left(F_{\mu'} - F_\mu\right)(x)\,\Phi(\mathrm{d}x) \\ &= \hat{C}(F_\mu) - \int \Phi(x)\left(F_{\mu'} - F_\mu\right)(\mathrm{d}x) \\ &= C(\mu) + \int (-\Phi)(x)\left(\mu' - \mu\right)(\mathrm{d}x), \end{aligned}$$

where the inequality follows from $\phi \in \partial \hat{C}\left(F_\mu\right)$, and the penultimate equality follows from integration by parts. Thus, we have

$$\begin{aligned}
\mu &\in \arg\max_{\mu' \in \mathcal{M}} \int \left(-\Phi\right)(x)\, \mu'\left(\mathrm{d}x\right) - C\left(\mu'\right) \\
&= \arg\max_{\mu' \in \mathcal{M}} \int \left(\max \Phi\left(X\right) - \Phi\right)(x)\, \mu'\left(\mathrm{d}x\right) - C\left(\mu'\right) \\
&= \arg\max_{\mu' \in \mathcal{M}} \int u\left(w\left(x\right)\right) \mu'\left(\mathrm{d}x\right) - C\left(\mu'\right),
\end{aligned}$$

as required.

■

Proof of Lemma 2. Observe first that, by Corollary 2, $W(\mu) = \int [w^*_\mu(x)]\mu(\mathrm{d}x)$. We begin by showing that

$$\mu \mapsto \int w^*_\mu(x)\mu(\mathrm{d}x) = \int u^{-1} \circ c_\mu(x)\mu(\mathrm{d}x)$$

is continuous. To this end, take any sequence $(\mu_n)_{n\in\mathbb{N}}$ that converges to some limit μ_∞. We first claim that

$$\lim_{n\to\infty} \|w^*_{\mu_n} - w^*_{\mu_\infty}\|_\infty = 0. \tag{11}$$

To prove this claim, fix some $\epsilon > 0$, take $T := [\min\ c_{\mu_\infty}(X) - \epsilon, \max\ c_{\mu_\infty}(X) + \epsilon]$, and let $S := u^{-1}(T) = \left[u^{-1}\left(\min\ c_{\mu_\infty}(X) - \epsilon\right), u^{-1}\left(\max\ c_{\mu_\infty}(X) + \epsilon\right)\right]$. Note that because u has a continuous and strictly positive derivative, $\bar{b} := \min_{s\in S}[u'(s)]$ is well defined and strictly positive, and so one can apply the Inverse Function Theorem to obtain that, for all $t \in T$, $\frac{\mathrm{d}u^{-1}}{\mathrm{d}t}(t) = \frac{1}{u'(u^{-1}(t))}$ is well-defined, strictly positive and bounded from above by $\bar{b}$. Therefore, the Mean Value Theorem implies that

$$|u^{-1}(t) - u^{-1}(t')| \leq \bar{b}|t - t'| \text{ for all } t, t' \in T. \tag{12}$$

To conclude the proof of the claim, fix some $\eta < \min\{\epsilon, \epsilon/\bar{b}\}$. By Assumption 3, an $N \in \mathbb{N}$ exists such that $\|c_{\mu_n} - c_{\mu_\infty}\|_\infty < \eta$ for all $n > N$. Therefore, all such n, $c_{\mu_n}(x)$ must be in T for all $x \in X$, and

$$\|w^*_{\mu_n} - w^*_{\mu_\infty}\|_\infty = \|u^{-1}(c_{\mu_n}) - u^1(c_{\mu_\infty})\|_\infty \leq \bar{b}\|c_{\mu_n} - c_{\mu_\infty}\| \leq \bar{b}\eta < \epsilon,$$

where the first inequality follows from (12), and the last from choice of η. Since ϵ was arbitrary, we have proven the claim that (11) holds. Armed with (11) one can prove

$\int w^*_\mu(x)\mu_n(\mathrm{d}x) \xrightarrow{n\to\infty} \int w^*_{\mu_\infty}(x)\mu_\infty(\mathrm{d}x)$ using the following inequality chain:

$$\begin{aligned}\left|\int w^*_{\mu_n}(x)\mu_n(\mathrm{d}x) - \int w^*_{\mu_\infty}(x)\mu_\infty(\mathrm{d}x)\right| &\leq \left|\int \left[w^*_{\mu_n}(x) - w^*_{\mu_\infty}(x)\right]\mu_\infty(\mathrm{d}x)\right| + \left|\int w^*_{\mu_\infty}(x)(\mu_n-\mu_\infty)(\mathrm{d}x)\right| \\ &\quad + \left|\int (w^*_{\mu_n} - w^*_{\mu_\infty})(x)(\mu_n-\mu_\infty)(\mathrm{d}x)\right| \\ &\leq \int \left|w^*_{\mu_n}(x) - w^*_{\mu_\infty}(x)\right|\mu_\infty(\mathrm{d}x) + \int \left|w^*_{\mu_\infty}(x)\right|(\mu_n-\mu_\infty)(\mathrm{d}x) \\ &\quad + \int \left|(w^*_{\mu_n} - w^*_{\mu_\infty})(x)\right|(\mu_n-\mu_\infty)(\mathrm{d}x) \\ &\xrightarrow{n\to\infty} 0,\end{aligned}$$

where convergence of the first and third term follow from (11), and convergence of the middle term following from $\mu_n \to \mu_\infty$ and $\left|w^*_{\mu_\infty}(\cdot)\right|$ being continuous.

Next, we prove that $\mu \mapsto \int w^*_\mu(x)\mu(\mathrm{d}x)$ is a Gateaux differentiable function admitting $w^*_\mu + \kappa^*_\mu(x)$ as its derivative. To this end, fix some $\tilde{\mu} \in \mathcal{M}$, and let $\mu_\epsilon = \mu + \epsilon(\tilde{\mu} - \mu)$. Observe

$$\begin{aligned}\frac{1}{\epsilon}\left[\int w^*_{\mu_\epsilon}(x)\,\mu_\epsilon(\mathrm{d}x) - \int w^*_\mu(x)\,\mu(\mathrm{d}x)\right] &= \int w^*_\mu(x)\,(\tilde{\mu}-\mu)(\mathrm{d}x) + \int \frac{1}{\epsilon}\left[w^*_{\mu_\epsilon} - w^*_\mu\right](x)\,\mu(\mathrm{d}x) \\ &\quad + \int \left[w^*_{\mu_\epsilon} - w^*_\mu\right](x)\,(\tilde{\mu}-\mu)(\mathrm{d}x).\end{aligned}$$

Hence, it is enough to show that

$$\lim_{\epsilon\searrow 0}\int \frac{1}{\epsilon}\left[w^*_{\mu_\epsilon} - w^*_\mu\right](x)\,\mu(\mathrm{d}x) = \int \kappa^*_\mu(y)\,(\tilde{\mu}-\mu)(\mathrm{d}y).$$

We now argue that, to show the above equality, it is sufficient to find a function $\phi : X \to \mathbb{R}$ that is integrable with respect to $(\tilde{\mu}-\mu)$, and an $\bar{\epsilon} \in (0,1)$ such that $\left|w^*_{\mu_\epsilon} - w^*_\mu\right| \leq \phi$ for all $\epsilon \in (0,\bar{\epsilon})$. To see why, note $\lim_{\epsilon\searrow 0}\left\|c_{\mu_\epsilon}(x) - c_\mu(x)\right\|_\infty = 0$ holds by Assumption 3, and so

$$\lim_{\epsilon\searrow 0}\left(\frac{u^{-1}\left(c_{\mu_\epsilon}(x)\right) - u^{-1}\left(c_\mu(x)\right)}{c_{\mu_\epsilon}(x) - c_\mu(x)}\right) = \frac{1}{u'\circ u^{-1}\left(c_\mu(x)\right)} = \frac{1}{u'\left(w^*_\mu(x)\right)}.$$

It follows that, for every x,

$$\begin{aligned}\lim_{\epsilon\searrow 0}\frac{1}{\epsilon}\left(w^*_{\mu_\epsilon}(x) - w^*_\mu(x)\right) &= \lim_{\epsilon\searrow 0}\frac{1}{\epsilon}\left(c_{\mu_\epsilon}(x) - c_\mu(x)\right)\left(\frac{u^{-1}\left(c_{\mu_\epsilon}(x)\right) - u^{-1}\left(c_\mu(x)\right)}{c_{\mu_\epsilon}(x) - c_\mu(x)}\right) \\ &= \int \frac{h_\mu(x,y)}{u'\left(w^*_\mu(x)\right)}(\tilde{\mu}-\mu)(\mathrm{d}y).\end{aligned}$$

Therefore, if a function ϕ as described above exists, the Lebesgue Dominated Convergence Theorem would imply that

$$\begin{aligned}\lim_{\epsilon \searrow 0} \int \frac{1}{\epsilon}\left[w^*_{\mu_\epsilon} - w^*_\mu\right](x)\,\mu(\mathrm{d}x) &= \int \lim_{\epsilon \searrow 0} \frac{1}{\epsilon}\left[w^*_{\mu_\epsilon} - w^*_\mu\right](x)\,\mu(\mathrm{d}x) \\ &= \int \frac{h_\mu(x,y)}{u'\left(w^*_\mu(x)\right)}(\tilde{\mu}-\mu)(\mathrm{d}y)\,\mu(\mathrm{d}x) \\ &= \int \frac{h_\mu(x,y)}{u'\left(w^*_\mu(x)\right)}\mu(\mathrm{d}x)(\tilde{\mu}-\mu)(\mathrm{d}y) = \int \kappa^*_\mu(y)(\tilde{\mu}-\mu)(\mathrm{d}y),\end{aligned}$$

as required.

We now find such a ϕ. Fix some $\eta > 0$, and note that Assumption 3 implies there is some $\bar{\epsilon} \in (0,1)$ such that for all $\epsilon \in (0,\bar{\epsilon})$ and all x,

$$\left|c_{\mu_\epsilon}(x) - c_\mu(x)\right| \leq \left|\int h_\mu(x,y)(\tilde{\mu}-\mu)(\mathrm{d}y)\right| + \eta.$$

Let

$$\bar{c} = \max_{x \in X}\left[c_\mu(x) + \left|\int h_\mu(x,y)(\tilde{\mu}-\mu)(\mathrm{d}y)\right|\right],$$

and take

$$b = \max_{y \in [0,\bar{c}+\eta]}\left(u^{-1}\right)'(y) = \max_{y \in [0,\bar{c}+\eta]} \frac{1}{u' \circ u^{-1}(y)},$$

which is finite and strictly positive, because u^{-1} is continuous and u' is strictly positive and continuous. Observe that, for every $\epsilon < \bar{\epsilon}$, and every x, the Mean Value Theorem implies there is some $a \in \mathrm{co}\left\{c_\mu(x), c_{\mu_\epsilon}(x)\right\} \subseteq [0, \bar{c}+\eta]$ such that

$$\frac{u^{-1}\left(c_{\mu_\epsilon}(x)\right) - u^{-1}\left(c_\mu(x)\right)}{c_{\mu_\epsilon}(x) - c_\mu(x)} = \left(u^{-1}\right)'(a) \leq b.$$

Therefore, for all $\epsilon < \bar{\epsilon}$ and every x,

$$\begin{aligned}\left|\frac{1}{\epsilon}\left(w^*_{\mu_\epsilon}(x) - w^*_\mu(x)\right)\right| &= \left|\frac{1}{\epsilon}\left(c_{\mu_\epsilon}(x) - c_\mu(x)\right)\left(\frac{u^{-1}\left(c_{\mu_\epsilon}(x)\right) - u^{-1}\left(c_\mu(x)\right)}{c_{\mu_\epsilon}(x) - c_\mu(x)}\right)\right| \\ &\leq \frac{1}{\epsilon}\left|c_{\mu_\epsilon}(x) - c_\mu(x)\right|\left|\frac{u^{-1}\left(c_{\mu_\epsilon}(x)\right) - u^{-1}\left(c_\mu(x)\right)}{c_{\mu_\epsilon}(x) - c_\mu(x)}\right| \\ &\leq \frac{b}{\epsilon}\left|c_{\mu_\epsilon}(x) - c_\mu(x)\right| \leq b\left(\int h_\mu(x,y)(\tilde{\mu}-\mu)(\mathrm{d}y)\right) + \eta.\end{aligned}$$

Thus, setting $\phi(x) = \eta + \int b h_\mu(x,y)(\tilde{\mu}-\mu)(\mathrm{d}y)$ gives the desired function. This concludes the proof. ■

Monotonicity of Gateaux Derivative. Next, we prove a result, required for generalizing Corollary 3.2 for the case in which Assumption 2 is relaxed to make c_μ lower semicontinuous (see footnotes 5 and 6 in the main text).

Lemma 3 *Suppose $C : \mathcal{M} \to \mathbb{R}$ is such that, for every μ there is a lower semicontinuous function $c_\mu : [0,1] \to \mathbb{R}$ for which $\int c_\mu(x)\ \mu(\mathrm{d}x)$ is finite and*

$$\lim_{\epsilon \downarrow 0} \frac{1}{\epsilon} \left[C(\mu + \epsilon(\mu' - \mu)) - C(\mu) \right] = \int c_\mu(x)\,(\mu' - \mu)\,(\mathrm{d}x)$$

for all $\mu' \in \mathcal{M}$. If C also satisfies Assumption 1, then c_μ is increasing.

Proof. Fix any $y, z \in X$ such that $y > z$. Observe that, for any $\epsilon \in [0,1]$, the distribution $\mu_{\epsilon,y} := \mu + \epsilon(\delta_y - \mu)$ first order dominates $\mu_{\epsilon,z} := \mu + \epsilon(\delta_z - \mu)$, where δ_y and δ_z are the distributions that respectively generate the outputs y and z for sure. We therefore obtain the following inequality chain:

$$\begin{aligned}
0 &\leq \frac{1}{\epsilon} \left[C(\mu_{\epsilon,y}) - C(\mu_{\epsilon,z}) \right] \\
&= \frac{1}{\epsilon} \left[C(\mu_{\epsilon,y}) - C(\mu) \right] + \frac{1}{\epsilon} \left[C(\mu) - C(\mu_{\epsilon,z}) \right] \\
&\xrightarrow{\epsilon \searrow 0} \int c_\mu(x)(\delta_y - \mu)(\mathrm{d}x) - \int c_\mu(x)(\delta_z - \mu)(\mathrm{d}x) \\
&= \int c_\mu(x)(\delta_y - \delta_z)(\mathrm{d}x) = c_\mu(y) - c_\mu(z).
\end{aligned}$$

The result follows. ■