# Contracting on what Firm Owners Value

Jonathan Bonham
University of Chicago
Booth School of Business

Amoray Riggs-Cragun
University of Chicago
Booth School of Business

July 26, 2021

**Abstract**

We revisit foundational questions in agency theory while assuming that the agent can fine-tune the joint distribution of all contractible and non-contractible performance measures. Under this assumption, optimal contracts behave as if the principal were making inferences about the outcome she values rather than about the agent's action. This has significant implications for what measures are included in contracts and how those measures are used. Most importantly, Holmström's (1979) informativeness principle changes. A performance measure is valuable if it improves inferences not about the agent's action, but about the outcome the principal values; and if that outcome is contractible, additional measures have no value. Our model predicts that contracts should be based on outcomes that firm owners intrinsically value, consistent with how real-world contracts tie executive pay to only a handful of accounting, market, or nonfinancial measures.

*Keywords:* Moral hazard, optimal contracts, performance measurement.

---

We appreciate helpful comments from Jeremy Bertomeu, Jörg Budde (discussant), Pierre Chaigneau, Alex Frankel, Robert Göx, Florian Heider, Thomas Hemmer, Bengt Holmström, Seung Lee, Christian Leuz, Valeri Nikolaev, Haresh Sapra, Abbie Smith, Lars Stole and participants in the DAR & DART Accounting Theory Seminar, the Chicago Booth Accounting Research Workshop, the Booth Junior Theory meeting, the Accounting and Economics Society webinar, and the 12th Accounting Research Workshop. We gratefully acknowledge financial support from the Booth School of Business and research assistance from Junyoung Jeong. The authors can be contacted at Jonathan.Bonham@chicagobooth.edu and amoray@chicagobooth.edu.

## 1 Introduction

Agency theory is a powerful tool for describing and analyzing conflicts of interest within organizations. However, some of its most fundamental predictions lack empirical support, particularly when it comes to executive compensation. The informativeness principle (Holmström 1979) says that optimal contracts should be conditioned on *all* costless signals containing incremental information about the agent's effort. Given that firms have enormous amounts of readily available information on hand, the informativeness principle would suggest that executive contracts should be conditioned on a large number of variables. However, recent evidence shows that the average compensation contract uses only three measures to determine a CEO's performance-based awards.[1] Furthermore, executive compensation contracts tend not to include performance measures that are informative about *effort* – number of hours spent in the office, meetings held, or emails sent – but instead seem to include measures that are informative about things firm owners value. For example, De Angelis and Grinstein (2015) document that 92 percent of performance-based awards are contingent on accounting metrics or stock performance. Overall, executive compensation contracts include fewer measures than classic agency models predict, and they are typically conditioned on information about things shareholders value, not managerial effort.

The above predictions are driven by one central insight from classic agency theory, that optimal contracts behave as if the principal were making inferences about the *agent's action*. In this paper, we return to Holmström's foundational moral hazard problem and change one assumption – what it means for the agent to take an action – and show that optimal contracts behave as if the principal were making inferences about *outcomes she values*. The aforementioned predictions (and perhaps many others) change as a result. In particular, our model predicts that contracts should be conditioned on measures of outcomes the principal intrinsically values rather than on measures of the agent's inputs.

Classical agency theory is grounded in the *parameterized distribution formulation* of the agency problem, in which the agent's action is represented as a parameter in the

[1] De Angelis and Grinstein 2015, Table 3, Panel C.

probability distribution over one or more random variables.[2] In this paper we invoke the *generalized distribution formulation*, wherein the agent directly and nonparametrically chooses the joint distribution over all relevant contractible and noncontractible performance measures.[3] This generalized approach expands the agent's choice set relative to the classic approach; rather than selecting the parameter of a distribution whose form is exogenous to his choice, the agent can implement *any* conceivable distribution (though some distributions are costlier to implement than others).

With intricate control over the distribution of performance, the agent in the generalized approach is far from an assembly line worker exerting unidimensional effort, but is better regarded as an appointed steward of the principal's assets. He is, in the fullest sense, an *agent*, one who acts in a decision-making capacity on another's behalf. This seems reasonably descriptive of the role played by corporate executives. By managing a company's product portfolio, investments, and strategic maneuvers, executives have significant influence not only on expected performance, but also on the *shape* of the distribution over performance. They can affect variance by managing the risk profile of the firm's project portfolio; they can increase skewness by making large R&D investments; and they can engage in activities that result in strange distributional shapes, such as discontinuities around zero earnings or at the point that just beats analyst forecasts.

We use the generalized distribution approach to revisit fundamental questions in agency theory. We begin with the univariate moral hazard problem from Holmström (1979), where the agent stochastically influences the principal's gross payoff, $x$ (which we term the "principal's objective"), and the principal writes a contract based on the realization of $x$. Under Holmström's parametric approach, the optimal sharing rule is shaped largely by the likelihood ratio. The contract behaves *as if* the principal

[2] The parameterized distribution formulation was first used by Mirrlees ([1975] 1999). Its use was widely popularized by Holmström (1979), who made the problem more tractable by assuming the first-order approach to be valid.

[3] The terms *parameterized distribution formulation* and *generalized distribution formulation* were coined by Hart and Holmström (1987). As we discuss in section 2, the generalized distribution formulation was devised by Holmström and Milgrom (1987) and was recently retooled and rigorously justified by Hébert (2018).

were using $x$ to make inferences about the agent's action, even though she knows what action the agent takes in equilibrium. As summarized by Murphy (1999), "The fundamental insight emerging from the traditional principal-agent models is that the optimal contract mimics a statistical inference problem: the payouts depend on the likelihood that the desired actions were in fact taken." Thus, in the classic approach, the contract depends on $x$ only through what $x$ says about the agent's action. We show that under the generalized distribution approach, likelihood ratios are absent in the solution – the optimal contract is conditioned on $x$ because it is valued by the principal, not because it is informative about the agent's action.

This result is driven by the relative dimensionality of the agent's and principal's opportunity sets, which is perhaps the most important difference between the classic parametric model and the generalized approach. In the classic model, the agent's action space is severely limited relative to the principal's contracting space, and as a result, there are infinitely many contracts that can implement a given action. As Holmström writes,

> "The optimum in the basic model tends to be complex, because there is an imbalance between the agent's one-dimensional effort space and the infinite-dimensional control space available to the principal. The principal is therefore able to fine-tune incentives by utilizing every minute piece of information about the agent's performance, resulting in complex schemes" (Holmström 2017).

The generalized distribution approach rectifies the control imbalance between the principal and the agent – the agent chooses the probability of each possible realization of $x$ and the principal chooses a contractual payment for each possible realization of $x$. Because the dimensionality of the two parties' opportunity sets are the same, the principal has limited options for controlling the agent; in fact, the implementing contract is unique. For a given contract, there is only one incentive compatible action, and for a given action that the principal would like to implement, there is only one contract that will do the trick.

Uniqueness also has significant implications for the value of conditioning the contract on some signal $y$ in addition to the objective $x$. Holmström's informativeness principle states that $y$ is valuable for contracting if and only if it provides incremental information about the agent's action. We show that this classic notion of informativeness does not hold under the generalized approach. When the agent chooses the joint distribution over $x$ and $y$ and both measures are contractible, the principal excludes $y$ from the contract. The reason for this, as before, stems from the equal balance of control between the principal and agent and, in turn, the uniqueness of the contract. In the classic model, the principal can use information in $y$ to improve contracting efficiency while holding the agent's action fixed. That is not possible in our setting. If the principal attempts to contract on $y$, the agent changes his actions to improve the distribution over $y$, and the principal must compensate the agent for this costly and unproductive behavior. In equilibrium, the contract is conditioned only on $x$ and the agent exerts no effort toward $y$. So long as the objective $x$ is contractible, it is the only measure used in the contract. This may help to explain why executive compensation contracts are conditioned on a small number of metrics.

But what if $x$ is not contractible? In this case, the control imbalance is tipped in the agent's favor: the agent chooses a joint distribution over both $x$ and $y$, while the principal chooses a sharing rule based only on $y$. Given the principal's control disadvantage, she has no choice but to induce costly manipulation of $y$ in order to make any improvements in $x$. We show that the contract that optimally manages this tradeoff is one that mimics a statistical inference problem: the optimal contract behaves as if the principal were using $y$ to estimate $x$. Therefore, while the classic model finds that $y$ is valuable if it is informative about the agent's action, we find that $y$ is valuable if it is informative about the realization of the objective $x$. This result may help to explain why executive compensation contracts tend to be conditioned on measures that are informative about firm value rather than on measures that are informative about the agent's inputs or effort.

When $x$ is not contractible, the solution can be separated into an estimation stage and a contracting stage. The principal first uses all available information to construct the best possible estimate of $x$ given the equilibrium distribution, and then she designs an optimal contract conditioned on that estimate. Thus, there is no loss to writing the contract on the estimate relative to writing it on all of the evidence that underlies the estimate. Therefore, even if $x$ is not contractible and there is a rich set of valuable signals available, the optimal contract may still be conditioned on only a handful of measures.

Our results can help to explain real-world compensation practices. The classic approach posits one reason for the inclusion of a given measure in a compensation contract: the measure is informative about the agent's actions. Taken together, our results suggest three alternative explanations. Specifically, the observation of a given measure in a compensation contract indicates that the measure is (i) something firm owners value, (ii) an estimate of something firm owners value, or (iii) an incrementally useful input for estimating something firm owners value. In section 6, we discuss the implications of our findings for the three major categories of measures used in CEO performance-based awards according to De Angelis and Grinstein (2015): accounting metrics, stock performance, and nonfinancial measures.

The paper proceeds as follows. In section 2, we discuss the origins of the generalized distribution approach, we explain how our paper is related to recent work using this approach, and we suggest that the generalized distribution approach might be interpreted as an abstraction of private information or multitasking models. In section 3, we introduce our model in a setting where the principal's objective, $x$, is the only contractible measure available. Here, we show that the optimal contract is unique and is conditioned on $x$ directly rather than through likelihood ratios. In section 4, we consider a setting in which the principal can contract on both $x$ and some set of additional measures, $\mathbf{y}$. We show that the agent's ability to game $\mathbf{y}$ results in the optimal contract conditioning only on $x$. In section 5, we show that when $x$ is not contractible, performance measures are valuable if they improve inferences about $x$; we provide a sufficient condition on the agent's cost function for a performance measure to be valuable for contracting; and we show that the solution can be decomposed into an estimation stage and a contracting stage. In section 6 we detail how our results help to explain common executive compensation practices. In section 7, we offer concluding remarks and suggest avenues for future work. In Appendix A, we show that our results can be generalized to broad classes of additively separable and multiplicatively separable cost functions. All proofs are provided in Appendix B.

## 2 Related Work

In a review of the early agency literature, Hart and Holmström (1987) identify three approaches to the moral hazard problem. The first is the *state space formulation* of early agency work (e.g. Wilson 1968), where output is jointly determined by the agent's action and a random state of nature. The second is the *parameterized distribution formulation* of the agency problem, wherein the agent's action is represented as a parameter in the distribution over one or more random variables. This approach has dominated the literature since the seminal work of Holmström (1979). Finally, Hart and Holmström describe the generalized approach as follows.

> The third, most abstract formulation is the following. Since the agent in effect chooses among alternative distributions, one is naturally led to take the distributions themselves as the actions, dropping the reference to $a$ ... Of course, the economic interpretation of the agent's action and the incurred cost is obscured in this *generalized distribution formulation*, but in return one gets a very streamlined model of particular use in understanding the formal structure of the problem. (Hart and Holmström 1987, pp.78-79.)

When Hart and Holmström (1987) wrote their review, the only paper to have used the generalized distribution approach was Holmström and Milgrom (1987).[4] Despite the benefits of this "streamlined approach," very few papers have used it since. This

[4]Specifically, Holmström and Milgrom (1987) use this approach in section 2 of their paper, where the agent is assumed to directly choose the probability of every outcome in a single-period model. Holmström and Milgrom (1987) are better known for what they do next. They divide the single period into multiple subperiods and show that when the subperiod length is taken to zero, the solution approximates a continuous time model in which the agent controls the drift of a a multi-dimensional Brownian process, and the optimal contract is linear in the ending positions of the process. This approximation is the birthplace of the widely-used "LEN" model, which exgoneously restricts the contract to be *L*inear, specifies that the agent has negative *E*xponential utility and assumes that the agent chooses the mean of a *N*ormal distribution. Note that the LEN specification falls under the classic parameterized

is perhaps because, as noted by Hart and Holmström in the quote above, interpreting the cost of the agent's action is not as straightforward in this approach as in the classic parametric model. A recent innovation is Hébert (2018), who microfounds the generalized (or *non-parametric*) approach using a cost function that is proportional to the divergence from some cost-minimizing distribution. We follow Hébert (2018) by using a divergence cost function as well.

Hébert (2018) studies optimal security design: a seller (the agent) has distributional control over an asset and designs a security that gives a buyer (the principal) some claim on the asset's value realization. (In our paper it is the principal who designs the contract, but this does not change the fundamental nature of the agency problem.) Our paper differs from Hébert (2018) in two important respects. First, we follow Holmström (1979) in modeling a risk-averse agent with unlimited liability, whereas Hébert assumes a risk-neutral agent with limited liability. Second, Hébert (2018) models a setting in which the principal's objective, $x$, is directly contractible and is the only measure considered in the contracting problem. While we also analyze that case (see section 3), much of our paper is focused on the contracting value of other performance measures when $x$ is available (section 4) and when $x$ is not available (section 5).

The other paper most closely related to ours is Bonham (2020), who uses the generalized approach to study how measurement and contracts shape productive incentives. The most important difference between the two papers has to do with what signals the agent can influence. In Bonham (2020), the agent has distributional control only over the principal's objective, $x$; the relationship between $x$ and some additional signal $y$ is completely exogenous to the agent's choice. In the present paper, the agent can influence the joint distribution over *all* performance measures, which produces different insights about the value of performance measures and how they are best used for contracting.

In addition to the papers mentioned above, we know of a few others that use

distribution formulation, because the agent chooses a parameter (the mean) of a normal distribution. Our adoption of the generalized distribution formulation from the single-period model in Holmström and Milgrom (1987) should not be confused with adoption of the LEN assumptions.

the generalized distribution approach. Hellwig (2007) uses it to extend Holmström and Milgrom (1987) to include boundary solutions. Bertomeu (2008) uses it to study risk management. Hemmer (2017) studies relative performance evaluation using a binary version of the generalized approach in which the agent directly chooses the probability of the principal's preferred outcome. Diamond (1998) takes a related but more restrictive approach in which the agent exerts costly effort to generate a set of distributions with equal means and then costlessly selects an element from that set.

The agent's rich action space is a key feature of the generalized approach, and we leverage it to study settings in which the dimensionality of the agent's opportunity set is at least as large as the principal's (a necessary ingredient for our results). Other papers, working within the classic parametric framework, have studied settings in which the agent has more power than he does in the classic single-action model. Although the agents in these models are still quite handicapped relative to one in the generalized approach, the forces at play are similar in flavor to the ones we study.

Multitasking models expand the agent's opportunity set by allowing the agent to choose a vector of effort levels, $\mathbf{a}$, directed toward various tasks that the principal may value differently. Notice that in this parametric setting, there is still a control imbalance that favors the principal. Even if the number of elements in $\mathbf{a}$ is taken to infinity, the agent has at most a *countably* infinite number of options, while the principal has *uncountably* many ways to structure the contract. As established in Holmström and Milgrom (1991), if the tasks in $\mathbf{a}$ differ in their measurability, incentives for the easy-to-measure tasks can induce unproductive effort allocations away from the hard-to-measure tasks. This is similar in spirit to the problem we study in section 5, where the principal cannot contract on the objective, $x$, and costly distortions arise from contracting on the available measure, $y$. However, it is the *agent* who has the control advantage in our setting, and our model produces very different solutions. In Holmström and Milgrom, if one of the tasks the principal values is unobservable, then the principal does best by offering *no* incentives for the measurable task; she offers a flat wage and the agent exerts some minimal amount of (positive) effort that he benevolently allocates according to the principal's preferences (Holmström and Milgrom 1991, Proposition 1). Thus, no costly distortions arise in equilibrium. In our setting, the *only* way to elicit effort towards productive but unmeasurable tasks, $x$, is by offering incentives for unproductive but measurable tasks, $y$, and consequently, the agent engages in some amount of costly distortive behavior in equilibrium.

Other papers have expanded the agent's opportunity set under the parametric approach by endowing the agent with private information. For example Raith (2008) studies a setting in which the agent has post-contractual private information about how his actions affect the principal's objective, $x$. To induce the agent to use his private information productively, the contract conditions on noisy measures of $x$, even if the agent's action is directly contractible. Baker (1992) studies a broader setting in which the agent has private information about the effect of his actions on both the objective $x$ as well as some additional measure, $y$. As in our section 5, $x$ is not contractible, and thus the principal is forced to write a contract on $y$ in order to incentivize any effort. The central tradeoff in Baker (1992) is that incentives on $y$ induce the agent to use his information about $x$ productively, but this comes at the cost of the agent using his information about $y$ unproductively. This is somewhat akin to the tradeoff at work in our section 5, where incentivizing improvements in $x$ comes at the cost of unproductive effort toward $y$.

Multitasking and private information models show that incentives can create distortions when the agent has multiple effort margins or privately known effort margins. The cost of distortive behavior is also an important force in our model, but our results are quite different from the aforementioned papers, and furthermore, our solutions are more general. Multitasking and private information models assume that contracts are linear; by contrast, the contracts in our model are endogenous and are thus optimal in the set of all possible contracts rather than just the linearly restricted ones.

The generalized distribution approach can be interpreted as a reduced form version of a broader setting in which the agent has an extremely rich action space. In this sense, it is not surprising that our model exhibits forces similar to those in multitasking and private information models. Suppose that an agent can engage in a wide array of activities, such as strategic planning, choosing investment projects, or managing the company's product mix. This agent could effectively choose the probability of every outcome by selecting an appropriate combination of actions from a sufficiently rich set of activities; we might then interpret the generalized distribution approach as a multitasking model taken to the extreme. Holmström and Milgrom

(1987) justify the generalized approach with several examples, including one in which the agent chooses a single action conditional on a rich set of private information in a static game; they argue that the space of contingent effort strategies maps to the space of nonparametric unconditional distributions at the outset, before any private information is revealed. Following this logic, the generalized approach might be interpreted as a very general private information model. Future research might show these connections formally, which would add to the work done by Hébert (2018) to micro-found the generalized distribution approach.

# 3 Revisiting the agency problem

In this section, we revisit the foundational univariate moral hazard setting wherein a principal contracts with an agent who takes an unobservable action that stochastically influences an outcome that the principal values, $x$. Following Baker (1992) and others, we term $x$ the principal's *objective*, whereas we refer to the principal's residual payoff (net of compensation paid to the agent) as her *objective function.* We do not take a stand on what it is that the principal cares about; $x$ may represent a project payoff, an asset value, stock price, cash flows, income, environmental impact, customer satisfaction or myriad other outcomes that a principal might intrinsically value (expressed in monetary terms).

We first present the classic formulation of the problem from Holmström (1979), in which the agent's action serves as a parameter in the probability distribution over $x$. We then recast the problem using the generalized distribution formulation, in which the agent chooses the distribution over $x$ nonparametrically.

## 3.1 The classic parameterized distribution approach

The classic formulation proceeds as follows. Let $x \in \mathbb{R}$ be the principal's objective. The agent takes a private action, $a \in A \subseteq \mathbb{R}$. This action, which is often interpreted simply as "effort," serves as a parameter in the probability density function $f(x; a)$, which we will take the liberty of calling the *distribution* over $x$. By his choice of $a$, the agent is effectively choosing a distribution from the set $\{f(x; a)|a \in A\}$. The agent incurs disutility of effort $V(a)$, where $V : A \to \mathbb{R}$, where $V(a)$ is increasing convex.

Because the agent dislikes working and acts privately, the principal must elicit effort from the agent using a performance-based inducement. Before the agent acts, the principal makes a one-shot offer of an incentive contract, $s$, which pays the agent $s(x)$ when the realized outcome is $x$. To simplify the exposition, we depart from Holmström (1979) by assuming that the principal is risk neutral such that her utility after paying the agent is given by $x - s(x)$. Let the agent have utility $H(s, a)$ and denote his utility from outside options as $\bar{H}$. Assume further that the agent's utility is additively separable in compensation and effort; specifically, let $H(s, a) = U(s(x)) - V(a)$. Finally, assume that $U'(\cdot) > 0$ and $U''(\cdot) < 0$; that is, the agent is strictly risk averse.

The principal's aim is to propose a contract and action pair $(s(\cdot), a)$ that maximizes her net payoff subject to two constraints. First, the proposed pair must make the agent's expected utility at least as great as $\bar{H}$; this individual rationality (IR) constraint ensures that the agent is willing to accept the contract. Second, the proposed pair must be incentive compatible (IC); that is, given the proposed scheme $s(x)$, the agent chooses the proposed $a$ voluntarily. Holmström (1979) simplifies incentive compatibility by assuming that the first-order approach is valid; that is, the agent's expected utility is globally concave in $a$ given $s(x)$. Given this assumption, the first-order condition from the agent's expected utility maximization problem is sufficient for $s(x)$ to be incentive compatible. With these classic ingredients in place, the principal's program is given as follows, where $f_a(x; a)$ denotes the derivative of $f(x; a)$ with respect to $a$.

$$\begin{aligned} \max_{s,a} \quad & \textstyle\int (x - s(x)) f(x; a) dx \\ \text{s.t.} \quad & \textstyle\int [U(s(x)) - V(a)] f(x; a) dx \geq \bar{H} \\ \text{and} \quad & \textstyle\int U(s(x)) f_a(x; a) dx = V'(a) \end{aligned} \tag{1}$$

Letting $\lambda$ and $\mu$ denote the Lagrange multipliers on the IR and IC constraints, pointwise optimization produces the following iconic characterization of the optimal sharing rule (Holmström 1979, equation 7).

$$\frac{1}{U'(s(x))} = \lambda + \mu \cdot \frac{f_a(x;a)}{f(x;a)} \tag{2}$$

The optimal contract depends critically on the likelihood ratio, $\frac{f_a(x;a)}{f(x;a)}$, which provides an indication of how likely it is that the agent took the proposed action given the outcome $x$. Thus, even though the principal knows that the agent will take the proposed action in equilibrium, the optimal contract behaves as if she were using the realization of $x$ to make inferences about the agent's action. This central insight from the parameterized approach permeates the agency literature.

### 3.2 Assumptions

We now invoke the generalized distribution formulation by assuming that rather than choosing a parametric distribution $f(x;a)$ from the set $\{f(x;a)|a \in A\}$, the agent has the ability to choose *any* $f(x)$ from the probability simplex $\Delta(x)$, which denotes the space of all probability distributions over $x$. We will assume implicitly that the agent's expected utility and cost function are computable under the chosen distribution.

As discussed in section 2, modeling the agent as choosing $f$ directly can be interpreted as the reduced form version of a broader setting in which the agent has an extremely rich action space. Holmström and Milgrom (1987) provide one example in which the agent acts continuously throughout the period, conditioning his action on a continuously observed state variable, and they argue that this setting can be represented in reduced form as the agent choosing the unconditional distribution at the outset. Hébert (2018) shows this formally, providing a micro-foundation for the generalized approach and also for a particular cost function, which we turn to now.

Let $g \in \Delta(x)$ be the agent's *preferred* or *cost-minimizing* distribution, the one that he would implement if offered zero incentives, and assume that $g$ has full support. Redefine the agent's personal cost as $V : \Delta(x) \to \mathbb{R}$, where $V(f)$ denotes the disutility incurred from implementing distribution $f$. Following Hébert (2018), we model $V(f)$

using the *Kullback-Leibler divergence* as follows.[5]

$$V(f) \quad = \quad D\left(f(x)||g(x)\right) \quad \equiv \quad \int f(x)\ln\left(\tfrac{f(x)}{g(x)}\right)dx. \tag{3}$$

KL divergence (or *relative entropy*) measures the dissimilarity between two distributions. This measure has been widely used in information theory and has many interpretations and applications.[6] For our purposes, it captures the cost incurred by the agent when he takes actions to implement some proposed distribution, $f$, rather than the cost-minimizing distribution, $g$. Notice from (3) that the cost of implementing $f(x)$ at $x$ is scaled by $g(x)$; intuitively, the agent dislikes large deviations from $g$ more than small ones.

Using the classic effort interpretation, we can think of $g$ as the distribution that is implemented when the agent exerts minimum effort.[7] This might be some ex ante distribution in place when the agent is hired, or alternatively, what the distribution over $x$ would degenerate to if the agent shirked his duties completely. Thinking beyond the effort interpretation, we could imagine $g$ as the "empire-building" distribution

[5]Hébert (2018) provides a micro-foundation for the use of a KL-divergence cost function under the generalized distribution approach. Specifically, he shows that a continuous time model in which an agent with quadratic cost controls the drift of a Brownian motion (akin to the continuous time model in Holmström and Milgrom 1987) is equivalent to a static model wherein an agent with KL cost chooses a probability distribution nonparametrically. We refer readers to Hébert (2018) for further justification of the KL-divergence cost function and the generalized approach.

[6]Kullback (1959) calls (3) the expected information for discrimination of $f$ in favor of $g$. In machine learning, KL divergence measures the loss of information that occurs if the reference distribution $f$ is approximated by distribution $g$. In Bayesian statistics, it reflects the information gained by moving from a prior distribution $g$ to a posterior distribution $f$.

[7]Hébert (2018) decomposes the total cost into a mean-shifting component, which he calls *effort*, and a *risk-shifting* component. While we could decompose the cost similarly, mean- versus risk-shifting is not our focus, and we will generally refer to any divergence from $g$ as requiring effort.

preferred by an agent seeking glory and acclaim at the expense of long-term value; or $g$ could represent the distribution over profits when the agent consumes lavish perquisites.

KL divergence has several properties that are relevant for our purposes. First, it is nonnegative. It is positive for all $f \neq g$, and it is zero when $f = g$; intuitively, the agent incurs zero personal cost when he implements his preferred distribution. Second, KL divergence is strictly convex in the pair $(f, g)$ if $f \neq g$ and is weakly convex if $f = g$ (Cover and Thomas 2006, Theorem 2.7.2). For a given $g$ with full support, it is increasingly costly for the agent to increase the probability of a given $x$, and it is maximally costly to implement a degenerate distribution that guarantees the realization of a particular $x$. Therefore, the agent's control over the *distribution* will almost never amount to full control over the *realization*, because the convexity of the cost function makes doing so maximally costly. In more applied terms, convexity implies that eliminating the firm's exposure to exogenous influences is practically infeasible. Third, the KL divergence marginal cost approaches infinity as $f(x)$ approaches zero for any $x$; this will simplify our analysis by guaranteeing interior solutions.

Finally, KL divergence is generally asymmetric; that is, the cost of moving the distribution from $g$ to $f$, $D\left(f(x)||g(x)\right)$, may be different from the cost of moving from $f$ to $g$, $D\left(g(x)||f(x)\right)$. This asymmetry fits naturally with many real-world production environments and human preferences. The costs of actions taken to enter a new market are likely to be different from those taken to leave a market; hiring someone is more fun than firing them; and the psychology of sunk costs can make it more painful to abandon a losing project than to start one in the first place.

We use KL divergence as our cost function due to its intuitive appeal and substantial foundations in information theory, but we do not need it to get our results. As we show in Appendix A, KL divergence is a special case of a much larger class of cost functions for which our results hold, including cost functions that are not additively separable in outcomes.

### 3.3 Analysis

We begin our analysis by examining the agent's choice problem after having signed the contract with the principal. Faced with contract $s$, the agent chooses $f$ to maximize his expected utility from compensation less his personal cost.

$$\begin{aligned} \max_{f} \quad & \int U(s(x))f(x)dx - \int f(x)\ln\left(\frac{f(x)}{g(x)}\right)dx \\ \text{s.t.} \quad & 1 = \int f(x)dx \\ & f(x) \geq 0 \text{ for all } x \end{aligned} \tag{4}$$

The constraints ensure that the chosen $f$ is a p.d.f. Let $\nu$ denote the Lagrange multiplier on the constraint $1 = \int f(x)dx$. We ignore the final set of constraints because, as we will verify presently, the KL cost function ensures that $f(x) \geq 0$ does not bind for any $x$ provided $g$ has full support. Pointwise optimization of (4) yields the following incentive compatible action.

$$f(x) \quad = \quad g(x)e^{U(s(x))-\nu-1}, \tag{5}$$

where $\nu = \ln\left(\int g(x)e^{U(s(x))-1}dx\right)$ is obtained by substituting (5) into the constraint $1 = \int f(x)dx$. Because both $g$ and the exponential function are positive everywhere, the unconstrained solution satisfies the constraint $f(x) \geq 0$ for all $x$. It follows that when faced with a particular incentive scheme $s$, the agent chooses $f$ such that (5) is maintained for all $x$. Let $w$ be some constant; then making the contract a flat wage by substituting $s(x) = w$ into (5) shows that $g$ is in fact the agent's preferred distribution, the one he implements when offered zero incentives.

$$f(x) \quad = \quad g(x)e^{U(w)-1-\ln\left(e^{U(w)-1}\int g(x)dx\right)} \quad = \quad g(x)e^{U(w)-1-(U(w)-1)} \quad = \quad g(x) \tag{6}$$

In the classic model, there are infinitely many contracts that satisfy the agent's first-order condition $\left(\int U(s(x))f_a(x;a)dx = V'(a)\right)$ and thus implement the desired $a$. As Holmström (2017) points out, this occurs because there is a severe control imbalance that favors the principal: the agent chooses the scalar $a$ while the principal chooses $s(x)$ at each $x$.Under the generalized approach, the agent chooses $f(x)$ at each $x$ and the principal chooses $s(x)$ at each $x$ – the opportunity sets of the two parties are of equal dimension. This equal balance of control produces a one-to-one mapping between the contract and the incentive-compatible action, as shown in the following lemma.

**Lemma 1** *For a given contract $s$, the incentive compatible distribution $f$ is unique. Moreover, for any distribution $f$ satisfying $f(x) > 0$ for all $x$, there exists an incentive compatible contract that is unique up to addition by a constant $\nu$ and is given in utility space by $U(s(x)) = \ln\left(\frac{f(x)}{g(x)}\right) + 1 + \nu$ for all $x$.*

Given a contract $s$, there is only one $f$ that solves the agent's problem, and there is only one incentive compatible $s$ that will implement a particular interior $f$ while awarding the agent a given level of utility. This is in contrast to the classic one-dimensional effort model, where the principal optimizes over infinitely many contracts that can implement a given action.

The first-order approach (FOA) is automatically valid under the generalized distribution approach. To see this, notice that program (4) maximizes a concave function with linear constraints.[8] We can therefore use the first-order conditions from (5) as the IC constraints in the principal's program. The principal's program is also constrained by $\nu = \ln\left(\int g(x)e^{U(s(x))-1}dx\right)$, which ensures that $\nu$ is specified such that $f(x)$ integrates to one in the agent's program, but it is straightforward to show that this constraint can be replaced by $1 = \int f(x)dx$.[9] Then the principal's program is written as follows, where for convenience we add and subtract $\nu$ to the left-hand side of the IR constraint.[10]

$$\begin{aligned} \max_{s,f} \quad & \int (x - s(x))f(x)dx \\ \text{s.t.} \quad & \nu + \int (U(s(x)) - \nu)f(x)dx - V(f) \geq \bar{H} \\ & U(s(x)) = \ln\left(\frac{f(x)}{g(x)}\right) + 1 + \nu \text{ for all } x \\ & 1 = \int f(x)dx \end{aligned} \tag{7}$$

[8]The Lagrangian is $\mathcal{L} = \nu + \int \left([U(s(x)) - \nu]f(x) - f(x)\ln\left(\frac{f(x)}{g(x)}\right)\right)dx$. By the additivity of integration, $\mathcal{L}$ is maximized over $f$ provided the function $\mathcal{L}(x) \equiv (U(s(x)) - \nu)\,f(x) - f(x)\ln\left(\frac{f(x)}{g(x)}\right)$ is maximized over $f(x)$ for all $x$. Because $(U(s(x)) - \nu)\,f(x)$ is linear in $f(x)$ and $f(x)\ln\left(\frac{f(x)}{g(x)}\right)$ convex in $f(x)$, $\mathcal{L}(x)$ is concave in $f(x)$.

[9]Substituting the incentive compatible contract $U(s(x)) = \ln\left(\frac{f(x)}{g(x)}\right) + \nu + 1$ into $\nu = \ln\left(\int g(x)e^{U(s(x))-1}dx\right)$ reduces it to $1 = \int f(x)dx$.

[10]An alternative way to write program (7) is to substitute $\nu = \ln\left(\int g(x)e^{U(s(x))-1}dx\right)$

Because (5) must be maintained at every $x$, there is an incentive compatibility constraint for each $x$. Throughout the paper, $\lambda$ and $\mu(x)$ denote the Lagrangian multipliers on the IR and IC constraints and $\eta$ denotes the multiplier on the constraint $1 = \int f(x)dx$. The following proposition characterizes the first-best solution to program (7).

**Proposition 1** *For purposes of establishing a first-best benchmark, assume that the agent is risk neutral so that $U(s(\cdot)) = s(\cdot)$. Then the optimal contract and action solving program* (7) *are characterized as follows.*

$$s(x) = x - \eta - 1 \tag{8}$$

$$f(x) = g(x)e^{x-\eta-1-\bar{H}}, \tag{9}$$

*where* $\eta = \ln\left(\int g(x)e^{x-\bar{H}-1}dx\right)$.

The principal sells the firm to the agent at price $\eta - 1$. While this is the standard first-best solution in moral hazard problems, here it is the *only* solution (due to the uniqueness result established in Lemma 1). This is in contrast to the classic approach, where *any* incentive-compatible contract that awards the agent his reservation utility is optimal when the agent is risk neutral.

into the IC constraint so that $\nu$ does not appear in the principal's program at all.

$$\begin{aligned} \max_{s,f} \quad & \int (x - s(x))f(x)dx \\ \text{s.t.} \quad & \int U(s(x))f(x)dx - V(f) \geq \bar{H} \\ & U(s(x)) = \ln\left(\tfrac{f(x)}{g(x)}\right) + 1 + \ln\left(\int g(\tilde{x})e^{U(s(\tilde{x}))-1}d\tilde{x}\right) \text{ for all } x \end{aligned}$$

It can be shown that the solution to this program is characterized by

$$\tfrac{1}{U'(s(x))} \quad = \quad \tilde{\lambda} + x - s(x),$$

where $\tilde{\lambda} \equiv \lambda + \lambda\nu - e^{\nu}\int \mu(\tilde{x})d\tilde{x}\left(\int g(\tilde{x})e^{U(s(\tilde{x}))-1}d\tilde{x}\right)^{-1}$. Although messier, this solution is of the same qualitative form as one we arrive at in Proposition 2; the right-hand side is equal to a constant plus $x - s(x)$.

Now we turn to the second-best solution. Pointwise optimization of (7) over $s(x)$ yields the following characterization of the optimal contract.

$$\frac{1}{U'(s(x))} = \lambda + \mu(x)\frac{1}{f(x)}. \tag{10}$$

Because the agent chooses each point in the distribution independently, the derivative of $f(x)$ with respect to $f$ at $x$ is equal to 1, so the likelihood ratio under the generalized approach is calculated as $\frac{\frac{d}{df}f(x)}{f(x)} = \frac{1}{f(x)}$. The term $\frac{1}{f(x)}$ is known in information theory as the *surprise* when $x$ is realized – the smaller the probability that a particular outcome $x$ will realize, the more surprised we are when it happens (Stone 2015, pp. 31–33). The term $\frac{1}{f(x)}$ in (10) is the direct analogue to the term $\frac{f_a(x;a)}{f(x;a)}$ from the classic characterization given by equation (2). The only substantial difference between the two contracts is that here there is a multiplier $\mu(x)$ for each $x$ (a result of uniqueness), whereas in Holmström (1979) there is only one such multiplier, $\mu$, because the principal only needs to ensure incentive compatibility for a single scalar, $a$.[11]

By the definition of a Lagrange multiplier, $\mu(x)$ captures the marginal degree to which ensuring incentive compatibility at $x$ reduces the principal's net payoff. Equivalently, we can think of $\mu(x)$ as indexing the severity of the conflict of interest between the principal and agent when it comes to the choice of $f$ at $x$. Equation (10) reveals that the shape of the optimal contract depends on both the conflict of interest, $\mu(x)$, as well as the likelihood ratio, $\frac{1}{f(x)}$; more specifically, it is the *interaction* between these components that determines the shape of the contract. Under the parameterized approach, the likelihood ratio runs the show; $\mu$ is constant in $x$ and so all variation in the contract comes through $\frac{f_a(x;a)}{f(x;a)}$. Under the generalized approach, the use of likelihood ratios is completely negated by the conflict of interest component;

[11] In the classic multitasking model, there are many multipliers as well – there is a $\mu_i$ for each $a_i \in \mathbf{a}$. In that framework, however, each first-order condition contains the entire vector of $\mu$'s, making the problem increasingly intractable as the dimensionality of $\mathbf{a}$ is increased. In our model, only one $\mu$ survives in each first-order condition; the only IC multiplier that appears in (10) is $\mu(x)$, which makes the problem much easier to work with.

uniqueness implies a separate $\mu(x)$ for each $x$, and the term $\mu(x)\frac{1}{f(x)}$ is replaced by the principal's objective function, as we show in the following proposition.[12]

**Proposition 2** *The optimal contract solving program* (7) *is characterized as follows.*

$$\frac{1}{U'(s(x))} = \tilde{\lambda} + x - s(x), \tag{11}$$

*where $\tilde{\lambda} = \lambda - \eta$. Moreover, $s(x)$ is increasing in $x$.*

While in the classic approach the contract depends on $x$ only through what $x$ says about the agent's action, the contract in (11) is conditioned *directly* on the principal's payoff, with no intermediate link to the likelihood that the agent took the the proposed action. Therefore, the central insight from classic agency theory – that contracts behave as if the principal were making inferences about the agent's action – is not robust to a generalized approach where the principal and agent have an equal balance of control.

## 4 The value of additional information when $x$ is contractible

Suppose that in addition to the objective $x$, the principal can contract on an $n$-dimensional vector of performance measures $\mathbf{y} \in \mathbb{R}^n$. The principal does not intrinsically care about $\mathbf{y}$ except through its impact on $x$. The vector $\mathbf{y}$ might include the number of projects undertaken, meetings held, hours worked, or emails sent (all of which are easily tracked on a company-issued laptop). Given a large amount of information available about the agent's activities, the question is whether and how this information should be incorporated into the agent's compensation contract.

[12] In Appendix A we show that KL divergence is a special case of a large class of cost functions wherein the use of likelihood ratios is completely negated by the conflict of interest component. Uniqueness is a necessary condition for this negation because it implies exactly one $\mu$ for each $x$; and for the class of cost functions identified in the appendix, uniqueness is both necessary and sufficient for the conflict of interest component to cancel out the likelihood ratio.

Holmström (1979) shows that when action $a$ parameterizes the joint distribution $f(x, \mathbf{y}; a)$, the optimal sharing rule, $s(x, \mathbf{y})$, is characterized as follows.

$$\frac{1}{U'(s(x,\mathbf{y}))} \quad = \quad \lambda + \mu \cdot \frac{f_a(x,\mathbf{y};a)}{f(x,\mathbf{y};a)} \tag{12}$$

Notice that $s(x, \mathbf{y})$ varies with $y_i$ to the extent that $\frac{f_a(x,\mathbf{y};a)}{f(x,\mathbf{y};a)}$ varies with $y_i$. If $\frac{f_a(x,\mathbf{y};a)}{f(x,\mathbf{y};a)}$ is constant in $y_i$, then $s(x, \mathbf{y})$ must be constant in $y_i$, which is equivalent to excluding $y_i$ from the contract altogether. Thus, the contract should be conditioned on some measure $y_i$ if the likelihood ratio varies in $y_i$. Holmström (1979) shows that this occurs if and only if $(x, \mathbf{y}_{-i})$ is not a sufficient statistic for $(x, \mathbf{y})$ with respect to $a$; that is, $y_i$ is valuable for contracting if it is incrementally informative about $a$. This is termed *the informativeness principle* and is considered one of the most robust results in agency theory (Bolton and Dewatripont 2005).

We now reassess the value of additional information using the generalized distribution approach. Assume that the agent directly chooses the joint distribution $f(x, \mathbf{y}) \in \Delta(x, \mathbf{y})$ at personal cost

$$V(f) \quad = \quad D(f(x,\mathbf{y})||g(x,\mathbf{y})) \quad \equiv \quad \int f(x,\mathbf{y}) \ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) d(x,\mathbf{y}), \tag{13}$$

where $\int d(x, \mathbf{y})$ indicates integration over $x$ and $\mathbf{y}$, $V : \Delta(x, \mathbf{y}) \to \mathbb{R}$, and $g(x, \mathbf{y}) \in \Delta(x, \mathbf{y})$ is the agent's preferred distribution absent incentives. This multivariate cost function has several economically appealing properties. First, it is additive in independent performance measures: if $x$ and $\mathbf{y}$ are independently distributed under both $f$ and $g$, then $D(f(x, \mathbf{y})||g(x, \mathbf{y})) = D(f(x)||g(x)) + D(f(\mathbf{y})||g(\mathbf{y}))$. Second, the agent's cost is weakly increasing in the number of performance measures he controls, so $D(f(x, \mathbf{y})||g(x, \mathbf{y})) \geq D(f(x)||g(x))$. Third, if the agent only exerts effort towards influencing $x$ while ignoring $\mathbf{y}$ so that $f(\mathbf{y}|x) = g(\mathbf{y}|x)$ for all $x$ and $\mathbf{y}$, then $D(f(x, \mathbf{y})||g(x, \mathbf{y})) = D(f(x)||g(x))$ (note that the converse is also true). In this case, $x$ is *sufficient for* $(x, \mathbf{y})$ *for the discrimination of* $f$ *from* $g$ (Kullback 1959 Theorem 4.1). We will elaborate on this property later on, as it will be useful for understanding our results.

Notice that there is still an equal balance of control between the principal and agent – the principal chooses the contractual payment $s(x, \mathbf{y})$ for each $(x, \mathbf{y})$ and the agent chooses $f(x,\mathbf{y})$ at each $(x,\mathbf{y})$. Uniqueness is therefore preserved in this setting, as stated formally in the following lemma. (The proof is identical to Lemma 1 except that $x$ is replaced everywhere by $(x,\mathbf{y})$).

**Lemma 2** *Let the agent choose the joint distribution $f(x,\mathbf{y})$, where both $x$ and $\mathbf{y}$ are contractible. For a given contract $s$, the incentive compatible distribution $f$ is unique. Moreover, for any distribution $f$ satisfying $f(x,\mathbf{y})>0$ for all $(x,\mathbf{y})$, there exists an incentive compatible contract that is unique up to addition by a constant $\nu$ and is given in utility space by $U(s(x,\mathbf{y}))=\ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right)+1+\nu$ for all $(x,\mathbf{y})$.*

Nonparametric control gives the agent an option that he does not have under the parametric approach: he can change $f(x)$ without changing $f(\mathbf{y}|x)$. In the classic approach, the agent chooses $a$ to parameterize $f(x,\mathbf{y};a)$, which by the definition of a joint probability density function can be decomposed as $f(x,\mathbf{y};a)=f(x;a)f(\mathbf{y}|x;a)$. The agent cannot change $f(x;a)$ independent of $f(\mathbf{y}|x;a)$ unless $f(\mathbf{y}|x;a)$ does not depend on $a$; that is, unless $x$ is a sufficient statistic for $(x,\mathbf{y})$ with respect to $a$. By contrast, in the generalized approach where the agent chooses $f(x,\mathbf{y})=f(\mathbf{y}|x)f(x)$ nonparametrically, he has the option to improve $f(x)$ without changing $f(\mathbf{y}|x)$.

To see why inducing this option may be attractive to the principal, we use the chain rule for relative entropy[13] to rewrite the cost function (13) as follows:

$$D\left(f(x,\mathbf{y})||g(x,\mathbf{y})\right) \;=\; D\left(f(x)||g(x)\right)+D\left(f(\mathbf{y}|x)||g(\mathbf{y}|x)\right), \tag{14}$$

where $D\left(f(\mathbf{y}|x)||g(\mathbf{y}|x)\right)\equiv\int\left(\int f(\mathbf{y}|x)\ln\left(\frac{f(\mathbf{y}|x)}{g(\mathbf{y}|x)}\right)d\mathbf{y}\right)f(x)dx$ is the conditional divergence (or *conditional relative entropy*) for $g(x,\mathbf{y})$ to $f(x,\mathbf{y})$. That is, the cost function can be decomposed into divergence in the marginal distribution from $g(x)$ to $f(x)$ plus the conditional divergence from $g(\mathbf{y}|x)$ to $f(\mathbf{y}|x)$. Conditional divergence is minimized and equal to zero when $f(\mathbf{y}|x)=g(\mathbf{y}|x)$ for all $x$. Therefore, the cheapest way for the agent to implement a given distribution $f(x)$ is to put no effort toward $f(\mathbf{y}|x)$. Because what the principal intrinsically cares about is $f(x)$ – not realizations of $\mathbf{y}$ or the conditional distribution $f(\mathbf{y}|x)$ *per se* – inducing $f(\mathbf{y}|x)=g(\mathbf{y}|x)$ for all $(x,\mathbf{y})$ appears to be an attractive option. The following proposition shows that if the

[13]See Cover and Thomas (2006), Theorem 2.5.3.

principal were to induce $f(\mathbf{y}|x) = g(\mathbf{y}|x)$, she could not condition the contract on $\mathbf{y}$.

**Proposition 3** *The agent implements $f(\mathbf{y}|x) = g(\mathbf{y}|x)$ for all $x$ if and only if $s(x, \mathbf{y})$ does not condition on $\mathbf{y}$.*

Due to the uniqueness result in Lemma 2, there is only *one* contract that can implement $f(\mathbf{y}|x) = g(\mathbf{y}|x)$ while holding $f(x)$ fixed. This is not the case under the classic approach, where a given level of effort $a$ can be implemented by infinitely many contracts $s(x, \mathbf{y})$, thereby allowing the principal to use the information in $\mathbf{y}$ to design a maximally efficient contract while holding $a$ fixed. By contrast, when the agent controls $f(x, \mathbf{y})$ nonparametrically and the dimensions of control are in balance, any attempt by the principal to use $\mathbf{y}$ results in the agent choosing an $f(\mathbf{y}|x)$ that diverges from $g(\mathbf{y}|x)$, and the principal must compensate the agent for this divergence. Thus, uniqueness imposes a cost to using $\mathbf{y}$ that is not present in the classic model: contracting on $\mathbf{y}$ induces unproductive changes in $f(\mathbf{y}|x)$. The following proposition shows that this cost outweighs any benefit from using $\mathbf{y}$; that is, it is indeed optimal for the principal to not contract on $\mathbf{y}$ and to thereby induce $f(\mathbf{y}|x) = g(\mathbf{y}|x)$.

**Proposition 4** *Let the agent choose $f(x, \mathbf{y})$, where the principal's objective is $x$ and $\mathbf{y}$ is some additional set of contractible performance measures. The optimal contract is characterized as follows.*

$$\tfrac{1}{U'(s(x,\mathbf{y}))} \;=\; \tilde{\lambda} + x - s(x, \mathbf{y}), \tag{15}$$

*where $\tilde{\lambda} = \lambda - \eta$. This expression can be maintained for all $\mathbf{y}$ only if $s(x, \mathbf{y})$ is increasing in $x$ and $s(x, \mathbf{y}) = s(x)$ for all $\mathbf{y}$. That is, the optimal contract does not condition on $\mathbf{y}$.*

Notice that the only place $\mathbf{y}$ appears in (15) is in $s(x, \mathbf{y})$; thus, if $s(x, \mathbf{y})$ varies in $\mathbf{y}$, equation (15) is violated for some $\mathbf{y}$. In stark contrast to the classic approach, if $x$ is available for contracting, the optimal contract excludes $\mathbf{y}$ completely.

Holmström's sufficient statistic condition says that $\mathbf{y}$ is valuable for contracting if it is incrementally informative about the agent's action. This classic version of informativeness does not hold under the generalized distribution approach because $\mathbf{y}$ does contain information about the action $f(x, \mathbf{y})$ in an absolute sense (the likelihood ratio $\frac{1}{f(x,\mathbf{y})}$ varies with $\mathbf{y}$) and yet the contract ignores $\mathbf{y}$. However, the condition that $f(\mathbf{y}|x) = g(\mathbf{y}|x)$ is the exact condition needed for $x$ to be *sufficient for* $(x, \mathbf{y})$ *for the discrimination of* $f$ *from* $g$ (see Kullback 1959, Chapter 2.4). It follows that a version of Holmström's informativeness principle is maintained here: the contract references signals that are incrementally informative about divergences from $g$ to $f$. In equilibrium, the *only* signal that satisfies this property is $x$.

**Corollary 1** *When the agent chooses* $f(x, \mathbf{y})$ *and* $x$ *is observable, the principal induces an equilibrium distribution that makes* $x$ *a sufficient statistic for* $(x, \mathbf{y})$ *for the discrimination of* $f$ *from* $g$.

Under the classic approach, the agent chooses $a$ to parameterize $f(x, \mathbf{y}; a) = f(x; a)f(\mathbf{y}|x; a)$, and $\mathbf{y}$ is incrementally informative about $a$ if $f(\mathbf{y}|x; a)$ varies in $a$; thus, the informativeness of $\mathbf{y}$ is exogenous. In our setting, the informativeness of $\mathbf{y}$ is determined endogenously – it is part of the solution that $x$ is sufficient for $(x, \mathbf{y})$ for the discrimination of $f$ from $g$. Thus, when the agent has nonparametric control over $f(x, \mathbf{y})$, scenarios in which $\mathbf{y}$ is incrementally informative are off the equilibrium path as long as $x$ is available for contracting.

## 5 The value of information when $x$ is not contractible

There may be settings in which the principal's objective, $x$, is not available for contracting. It is common practice in accounting theory to assume that the "firm fundamentals" shareholders care about are unobservable, whereas observable accounting metrics serve as signals about the underlying fundamentals. In this section, we examine the contracting value of available signals, $\mathbf{y}$, when the objective, $x$, is not contractible.

Holmström (1979) does not explicitly consider the case in which $x$ is not contractible, but extending his approach to this setting is straightforward. If $a$ parameterizes $f(x, \mathbf{y}; a)$ as before but only $\mathbf{y}$ is contractible, pointwise optimization with respect to $s(\mathbf{y})$ yields the following.

$$\frac{1}{U'(s(\mathbf{y}))} = \lambda + \mu \cdot \frac{f_a(\mathbf{y}; a)}{f(\mathbf{y}; a)} \tag{16}$$

The central insight from the original setting is preserved: the optimal contract behaves as if the principal were using $\mathbf{y}$ to make inferences about $a$, and therefore, signals in $\mathbf{y}$ are included in the contract if they are incrementally informative about $a$. The reason that making $x$ non-contractible has no effect on the classic result is that doing so does not meaningfully change the control imbalance between the principal and the agent – the agent has a single dimension of control, $a$, while the principal still has infinite dimension in her choice of $s(\mathbf{y})$.

Under the generalized distribution approach, making $x$ non-contractible tips the balance of control in the agent's favor. The agent chooses $f(x, \mathbf{y})$ for every $(x, \mathbf{y})$ while the principal chooses $s(\mathbf{y})$ for every $\mathbf{y}$. For the case where $\mathbf{y}$ is a scalar, we can envision the agent as choosing a surface while the principal chooses only a curve. Because the principal has relatively limited control, there will be some distributions that she cannot induce with any contract. However, a weaker form of uniqueness is preserved: within the set of distributions that can be implemented, the implementing contract is unique.

**Lemma 3** *Let the agent choose the joint distribution $f(x, \mathbf{y})$, where the elements of $\mathbf{y}$ are contractible but the objective $x$ is not. For a given contract $s$, the incentive compatible distribution $f$ is unique and satisfies $f(x, \mathbf{y}) > 0$ and $f(x|\mathbf{y}) = g(x|\mathbf{y})$ for all $(x, \mathbf{y})$. Any distribution satisfying $f(x|\mathbf{y}) \neq g(x|\mathbf{y})$ for some $(x, \mathbf{y})$ cannot be implemented by any contract written only on $\mathbf{y}$. Moreover, for any distribution $f$ satisfying $f(x, \mathbf{y}) > 0$ and $f(x|\mathbf{y}) = g(x|\mathbf{y})$, there exists an incentive compatible contract that is unique up to addition by a constant $\nu$ and is given in utility space by $U(s(\mathbf{y})) = \ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) + \nu + 1 = \ln\left(\frac{f(\mathbf{y})}{g(\mathbf{y})}\right) + \nu + 1$ for all $\mathbf{y}$.*

The agent's control advantage prevents the principal from implementing any conditional distribution other than $f(x|\mathbf{y}) = g(x|\mathbf{y})$. Thus, the principal's problem is to write a contract on $\mathbf{y}$ that results in the most productive $g(x|\mathbf{y})$ possible. The solution to the principal's problem is provided in the following proposition.

**Proposition 5** *If the principal cannot contract on the objective, $x$, but can contract on some other set of signals, $\mathbf{y}$, then the optimal contract is characterized by the* *following expression.*

$$\frac{1}{U'(s(\mathbf{y}))} \quad = \quad \tilde{\lambda} + \mathbb{E}_f[x|\mathbf{y}] - s(\mathbf{y}), \tag{17}$$

*where* $\tilde{\lambda} = \lambda - \eta$ *and* $\mathbb{E}_f[x|\mathbf{y}]$ *is the expected value of* $x$ *given* $\mathbf{y}$ *under the agent's equilibrium action* $f$*. The expression can be maintained for all* $\mathbf{y}$ *only if* $s(\mathbf{y})$ *is increasing in* $\mathbb{E}_f[x|\mathbf{y}]$*. In particular,* $s(\mathbf{y})$ *varies with* $y_i$ *if and only if* $\mathbb{E}_f[x|\mathbf{y}]$ *varies with* $y_i$*, i.e., if* $y_i$ *contains incremental information about* $x$ *(as opposed to the action* $f$*) in equilibrium.*

The proposition shows that the optimal contract behaves as if the principal were using information in $\mathbf{y}$ make inferences about *the principal's objective*, $x$, given the equilibrium distribution. It follows that performance measures are valuable for contracting if they improve inferences about the principal's objective. This is in contrast to the classic model, where performance measures are valuable if they improve inferences about the *agent's action.* The following proposition provides a necessary and sufficient condition for a signal to improve inferences about $x$ and therefore to have contracting value.

**Proposition 6** *Let the agent choose the joint distribution* $f(x, \mathbf{y})$*, where the elements of* $\mathbf{y}$ *are contractible but the objective* $x$ *is not. Some signal* $y_i \in \mathbf{y}$ *is valuable for contracting if and only if* $E_g[x|\mathbf{y}] \neq E_g[x|\mathbf{y}_{-i}]$ *for some realization of* $\mathbf{y}$*.*

The proposition shows that the optimal contract is conditioned on some signal $y_i$ if and only if $E_g[x|\mathbf{y}]$ varies in $y_i$. That is, for $y_i$ to have contracting value, it must be informative about the expected value of $x$ under the agent's preferred distribution $g$. As an example, assume that the agent is a salesperson and that $x$ is the likelihood that a client will continue his relationship with the company. If the agent enjoys golf and is inclined to talk a little business while on the course, then it might be productive to reward the him for the number of rounds he plays with the client. If on the other hand being on the golf course does not change the agent's willingness to sell, then the number of golf rounds played has no contracting value.

Section 4 showed that when $x$ is contractible, the optimal contract conditions on $x$ and ignores all other signals. This seems consistent with real-world contracts, which include far fewer measures than predicted by the classic approach. It may appear from Propositions 5 and 6 that when $x$ is not contractible, optimal contracts will again be conditioned on a large number of signals. However, inspection of (17) shows that the solution can be decomposed into two steps. First, the principal uses information in $\mathbf{y}$ to construct an unbiased estimate of $x$; denote this estimate by $\hat{x}$. Second, the principal writes a contract on $\hat{x}$. This is formalized in the following corollary.

**Corollary 2** *Let the agent choose $f(x, \mathbf{y})$, the joint distribution over the objective $x$ and some other set of measures, $\mathbf{y}$. If $\mathbf{y}$ is contractible but $x$ is not, then an optimal estimator and contract solving the principal's program are characterized as follows.*

$$\begin{aligned} \hat{x} &= \mathbb{E}_f[x|\mathbf{y}] \\ s(\hat{x}) &= \tilde{U}^{-1}(\hat{x} + \tilde{\lambda}), \end{aligned} \tag{18}$$

*where $\tilde{U} = s + \frac{1}{U'(s)}$ and $\tilde{\lambda} = \lambda - \eta$. In particular, there is no loss to contracting on the estimate $\hat{x}$ relative to contracting on $\mathbf{y}$.*

The corollary shows that using signals in the contract is equivalent to using signals in estimating $x$. Therefore, even when $x$ is not contractible and there is a rich set of contractible performance measures available, these measures may not appear explicitly in the optimal contract. Instead, the contract may condition on just a few estimates of things the principal values.

## 6 Empirical Implications

We have shown an approach to agency theory in which optimal contracts behave as if the principal were making inferences about outcomes rather than actions. In this section, we suggest a few ways that this finding can speak to three types of performance measures commonly used in practice: accounting metrics, stock price, and nonfinancial performance measures.

**Accounting-based performance measures**

Companies have at their disposal enormous databases of disaggregated accounting information, such as the timing and amounts of all transactions with customers and suppliers. Executive compensation contracts are not directly conditioned on all of this data; De Angelis and Grinstein (2015) document that seventy-nine percent of performance-based awards in CEO compensation contracts are based on accounting metrics and that the majority of those metrics are earnings measures (such as net income, EBIT, or EBITDA). This is not easily explained in the classic approach. Holmström's informativeness principal suggests that contracts should depend on all disaggregated signals unless there exists an aggregate that is a sufficient statistic with respect to the agent's action. It seems unlikely that earnings could be a sufficient statistic for all of the underlying data used to construct it. Rationalizing the use of estimates like net income under the classic approach thus requires imposing some friction that makes contracting on disaggregated signals infeasible.

Our model provides a theoretical explanation for contracting on accounting estimates. Let $\mathbf{y}$ represent detailed information such individual sales data; inventory delivery times; the quantity and prices of inventory purchases; the amounts and likely success of various R&D projects; and the historical purchase prices and replacement costs of all fixed assets in place. Assume that shareholders care about *true earnings*, the amount of real value created during the contracting period, and denote true earnings by $x$. Proposition 5 shows that the optimal contract behaves as if the principal were using the information in $\mathbf{y}$ to make inferences about true earnings, $x$. Corollary 2 shows that this solution can be broken into two stages: *estimation* and *contracting*. The principal first uses the information in $\mathbf{y}$ to construct an unbiased estimate of true earnings, denoted $\hat{x}$, and then she writes a contract that ties the agent's compensation directly to this estimate. Thus, to the extent that accounting net income is an unbiased estimate of true earnings, there is no loss to contracting directly on net income relative to contracting on all of the underlying information used to construct net income.

**Stock price as a performance measure**

Prior papers find that it is inefficient to condition compensation contracts on stock price alone because information is aggregated differently for stewardship (i.e. contracting) and valuation purposes (Paul 1992; Feltham and Xie 1994; Lambert 2001).[14] Our results provide a benchmark in which there is no such conflict. Let $x$ be true firm value, and suppose that $\mathbf{y}$ represents a set of signals that is available for both valuing the firm and compensating the agent. Firm stock price, $p(\mathbf{y}) \equiv \mathbb{E}_f[x|\mathbf{y}]$, uses information in $\mathbf{y}$ to construct an estimate of true firm value, $x$, given the agent's equilibrium choice of $f(x, \mathbf{y})$. By Corollary 2, the optimal contract is conditioned on $\hat{x} = \mathbb{E}_f[x|\mathbf{y}]$. Then $\hat{x} = p(\mathbf{y})$, and therefore there is no loss to contracting on stock price alone relative to contracting on $\mathbf{y}$.

There is also no conflict between the stewardship and valuation roles of information under our approach when stock price itself is the principal's objective. If $x$ is stock price, it follows immediately from Proposition 4 that the principal will condition the contract exclusively on stock price and ignore the information in $\mathbf{y}$. This is in stark contrast to the classic parametric approach, where the optimal contract will include stock price as well as any additional signals in $\mathbf{y}$ that are incrementally informative about the agent's action.[15]

It is possible that stock prices do a poor job of constructing an unbiased estimate of true firm value, but that shareholders care to some extent about stock price nonetheless. We do not take a stand on what variables enter principal's objective, what information is impounded into price or whether stock markets are efficient. Instead, these examples are meant to suggest some context for our more general re-

[14]To see how this conclusion is reached in the classic setting, suppose that the principal's objective is given by $x = b_1 a_1 + b_2 a_2 + \epsilon_x$ and that there are two performance measures available, $y_i = b_i a_i + \epsilon_i$, $i = \{1, 2\}$. Then the pricing function function assigns $y_i$ a weight of $\frac{\text{cov}(x, y_i)}{\text{var}(y_i)}$, while the compensation contract assigns $y_i$ a weight of $\frac{b_i^2}{b_i^2 + r\text{var}(y_i)}$. Thus, there is no loss to contracting on price alone if and only if these two weights are proportional for all $i$, which is unlikely to hold unless by coincidence.

[15]See section 3.3.5 of Lambert (2001) for a detailed discussion of why Holmström's informativeness principle prevents stock price from being sufficient for contracting.

sult, which is that it is efficient to contract on estimates of the principal's objective, whatever that objective may be. We leave it to future research to conduct more fully-developed analyses of contracting on specific measures such as stock price.

**Nonfinancial performance measures**

Recent evidence shows a growing trend in tying executive compensation to corporate social responsibility (CSR) initiatives. Ikram, Li, and Minor (2019) examine proxy statements and find that the percentage of firms linking executive compensation to CSR measures increased monotonically from 42.5% in 2009 to 47.6% in 2013. These authors (and others) connect the practice of contracting on CSR initiatives to agency theory through the informativeness principle, suggesting that CSR contracts provide additional information about managerial effort. Our results suggest alternative agency theoretic explanations for the use of CSR in executive compensation contracts. Specifically, our results suggest that a contract is conditioned on a CSR measure for one of two reasons: either (i) shareholders intrinsically care about CSR, or (ii) the CSR measure is useful for estimating something that shareholders care about. To illustrate the second reason, suppose that shareholders care only about true firm value. It is possible that CSR initiatives have reputational effects, which could impact future profits. In this case, CSR measures may enter the contract because they provide sharper inferences about the principal's objective, true firm value.

There is empirical evidence that other nonfinancial performance measures are used in CEO compensation contracts as well. Using a sample of airline firms, Davila and Venkatachalam (2004) find that passenger load factor – the total number of miles flown by passengers divided by the number of seat-miles available – is positively associated with CEO cash compensation after controlling for financial performance. They state that this evidence is consistent with the classic informativeness principal, writing that passenger load factor provides "incremental information about CEO's actions over financial measures." Our results suggest an alternative explanation: passenger load factor may sharpen inferences about firm value.

## 7 Conclusion

We use the generalized distribution approach to revisit optimal contracting and the value of information under moral hazard. In stark contrast to the bulk of the agency literature, we find that performance measures are valuable when they are informative about the principal's objective, not the agent's action. Our results are driven by the balance of control between the principal and agent under the generalized approach – the agent has at least as much control over performance as the principal has over the agent. This assumption reflects the separation of ownership and control inherent in publicly-held corporations (e.g. Fama and Jensen 1983).

This paper opens multiple avenues for future research. First, the model could be expanded to include multiple agents or multiple periods. Such analyses are likely to be fruitful because the informativeness principle permeates the findings from these extensions in the classic model. Second, our results may provide opportunities to revisit empirical studies of executive compensation, as the empirical hypotheses in those papers are almost always derived from the classic notion of informativeness. Third, while our results hold for a large class of cost functions (see Appendix A), it is possible to write down cost functions such that uniqueness is not sufficient to guarantee the negation of likelihood ratios. Future research could examine whether such cost functions have economically defensible properties. Finally, it would be useful if future research could provide a reconciliation between the generalized approach and the parametric approach, which would aid in comparing findings across the two approaches.

## References

Baker, George P. 1992. Incentive contracts and performance measurement. *Journal of Political Economy* 100 (3): 598–614.

Bertomeu, Jeremy. 2008. Risk management and the design of efficient incentive schemes. PhD diss., Carnegie Mellon University.

Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract theory.* MIT press.

Bonham, Jonathan. 2020. Shaping incentives through measurement and contracts. Working Paper.

Cover, Thomas M, and Joy A. Thomas. 2006. *Elements of information theory, 2nd edition.* John Wiley & Sons, Inc.

Davila, Antonio, and Mohan Venkatachalam. 2004. The relevance of non-financial performance measures for CEO compensation. *Review of Accounting Studies* 9 (4): 443–464.

De Angelis, David, and Yaniv Grinstein. 2015. Performance terms in CEO compensation contracts. *Review of Finance* 19 (2): 619–651.

Diamond, Peter. 1998. Managerial incentives: On the near linearity of optimal compensation. *Journal of Political Economy* 106 (5): 931–957.

Fama, Eugene F, and Michael C Jensen. 1983. Agency problems and residual claims. *The Journal of Law and Economics* 26 (2): 327–349.

Feltham, Gerald A, and Jim Xie. 1994. Performance measure congruity and diversity in multi-task principal/agent relations. *The Accounting Review* 69 (3): 429–453.

Hart, Oliver, and Bengt Holmström. 1987. The theory of contracts. In *Advances in Economic Theory: Fifth World Congress,* vol. 1. Cambridge University Press.

Hébert, Benjamin. 2018. Moral hazard and the optimality of debt. *The Review of Economic Studies* 85 (4): 2214–2252.

Hellwig, Martin F. 2007. The role of boundary solutions in principal–agent problems of the Holmström–Milgrom type. *Journal of Economic Theory* 136 (1): 446–475.

Hemmer, Thomas. 2017. Optimal dynamic relative performance evaluation. Working Paper.

Holmström, Bengt. 1979. Moral hazard and observability. *The Bell Journal of Economics* 10 (1): 74–91.

———. 2017. Pay for performance and beyond. *American Economic Review* 107 (7): 1753–77.

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

———. 1991. Multitask principal-agent analyses: incentive contracts, asset ownership, and job design. *Journal of Law, Economics and Organization* 7:24–52.

Ikram, Atif, Zhichuan Frank Li, and Dylan Minor. In press. CSR-contingent executive compensation contracts. *Journal of Banking & Finance.*

Kullback, Solomon. 1959. *Information theory and statistics.* John Wiley & Sons.

Lambert, Richard A. 2001. Contracting theory and accounting. *Journal of Accounting and Economics* 32 (1-3): 3–87.

Mirrlees, James. (1975) 1999. The theory of moral hazard and unobservable behaviour. Part I. *The Review of Economic Studies* 66 (1): 3–21. Paper completed 1975.

Murphy, Kevin J. 1999. Executive compensation. *Handbook of Labor Economics* 3.

Paul, Jonathan M. 1992. On the efficiency of stock-based compensation. *The Review of Financial Studies* 5 (3): 471–502.

Raith, Michael. 2008. Specific knowledge and performance measurement. *The RAND Journal of Economics* 39 (4): 1059–1079.

Stone, James V. 2015. *Information theory: A tutorial introduction.* Sebtel Press.

Wilson, Robert. 1968. The theory of syndicates. *Econometrica* 36 (1): 119–132.

# A Cost Function Robustness

We have used the KL-divergence cost function throughout the paper because it has well-established foundations in information theory. However, our results are robust to a much larger class of cost functions. In this appendix, we establish sufficient conditions for our results to hold when $V(f)$ is additively separable (section A.1) and multiplicatively separable (section A.2), and we provide examples throughout of cost functions that satisfy the sufficient conditions.

## A.1 Additive separability

Let the agent control $f(x)$ at personal cost $V(f)$, given as follows.

$$V(f) \quad = \quad \int v_x(f(x))dx, \tag{19}$$

where for each $x$, the function $v_x : \mathbb{R} \to \mathbb{R}$ gives the cost of implementing $f(x)$ at $x$. The cost $V(f)$ is additively separable across $x$. Additive separability allows for the possibility that there is a different function $v$ for each $x$, which is why we index $v_x$ by $x$. For KL-divergence, $v_x(f(x)) = f(x) \ln \left(\frac{f(x)}{g(x)}\right)$ for all $x$. Let $v'_x(f(x))$ denote the first derivative of $v_x$ with respect to $f$, and let $v''_x(f(x))$ denote the second derivative of $v_x$ with respect to $f$. Assume that the second derivative is positive; that is, the agent's cost of increasing $f$ at $x$ is convex.

Faced with contract $s$, the agent chooses $f$ to maximize his expected utility from compensation less his personal cost.

$$\begin{aligned} \max_f \quad & \int U(s(x))f(x)dx - \int v_x(f(x))dx \\ \text{s.t.} \quad & 1 = \int f(x)dx \end{aligned} \tag{20}$$

Let $\nu$ denote the Lagrangian multiplier on the constraint $1 = \int f(x)dx$. Then pointwise optimization yields the following characterization of the incentive compatible action.

$$v'_x(f(x)) \quad = \quad U(s(x)) - \nu. \tag{21}$$

The principal's program is given as follows, where for convenience we add and subtract $\nu$ to the left-hand side of the IR constraint.

$$\begin{aligned} \max_{s,f} \quad & \int (x - s(x)) f(x) dx \\ \text{s.t.} \quad & \nu + \int (U(s(x)) - \nu) f(x) dx - \int v_x(f(x)) dx \geq \bar{H} \\ & U(s(x)) = \nu + v_x'(f(x)) \text{ for all } x \\ & 1 = \int f(x) dx \end{aligned} \tag{22}$$

Pointwise optimization yields the following characterization of the optimal contract.

$$\frac{1}{U'(s(x))} \quad = \quad \lambda + \mu(x) \frac{1}{f(x)}. \tag{23}$$

Optimizing (22) pointwise with respect to $f$ allows us to solve for $\mu(x)$ in closed form as $\mu(x) = \frac{x - s(x) - \eta}{v_x''(f(x))}$. Substituting this expression into (23) gives

$$\frac{1}{U'(s(x))} \quad = \quad \lambda + \left( \frac{x - s(x) - \eta}{v_x''(f(x))} \right) \left( \frac{1}{f(x)} \right). \tag{24}$$

Suppose that $v_x(f(x))$ is specified such that the product of $v_x''(f(x))$ and $f(x)$ can be written as a transformation of the first derivative, $v_x'(f(x))$. Then because the incentive compatible action characterized in (21) gives us that $v_x'(f(x)) = U(s(x)) - \nu$, we can replace the denominator $v_x''(f(x))f(x)$ in (24) with a transformation of $U(s(x)) - \nu$. This is formalized in the following proposition.

**Proposition A.1** *Let the agent choose the distribution over $x$ at personal cost $V(f) = \int v_x(f(x))dx$. If $V(f)$ admits the existence of a function $\tilde{v}$ such that $v_x''(f(x))f(x) = \tilde{v}(v_x'(f(x)))$ for all $x$, where $x$ enters the argument of $\tilde{v}$ only through $v_x'(f(x))$, then the optimal contract solving program (22) is characterized as follows.*

$$\frac{1}{U'(s(x))} \quad = \quad \lambda + \frac{x - s(x) - \eta}{\tilde{v}(U(s(x)) - \nu)} \tag{25}$$

*Moreover, if $\tilde{v}$ is nondecreasing, then $s(x)$ is increasing in $x$.*

The proposition shows that the finding from Proposition 2 is preserved for all additively separable cost functions for which the function $\tilde{v}$ exists. In particular, the likelihood ratio does not appear in the contract and neither does the agent's choice of $f(x)$.

To see that the condition in Proposition A.1 is satisfied by the KL divergence cost function, notice that if $v_x(f(x)) = f(x)\ln\left(\frac{f(x)}{g(x)}\right)$ then $v_x'(f(x)) = \ln\left(\frac{f(x)}{g(x)}\right) + 1$ and $v_x''(f(x)) = \frac{1}{f(x)}$, so $v_x''(f(x))f(x) = 1$. Then the function $\tilde{v}$ exists and is given by $\tilde{v}(v_x'(f(x))) = 1$.

The condition in Proposition A.1 is satisfied for a large class of cost functions. For example, it is satisfied for all cost functions for which $v_x'(f(x))$ is homogeneous of any degree. To see this, note that if $v_x'(f(x))$ is homogeneous of degree $k$ so that $v_x'(\alpha f(x)) = \alpha^k v_x'(f(x))$, then Euler's theorem for homogeneous functions gives that $f(x)v_x''(f(x)) = kv_x'(f(x))$. This includes the entire power class. As an example, assume that the agent has the quadratic cost function $V(f) = \int \frac{f(x)^2}{2g(x)}dx$. Then $v_x'(f(x)) = \frac{f(x)}{g(x)}$ and $v_x''(f(x)) = \frac{1}{g(x)}$. Multiplying $v_x''(f(x))$ and $f(x)$ recovers the first derivative: $v_x''(f(x))f(x) = \frac{1}{g(x)}f(x) = v_x'(f(x))$, implying that $\tilde{v}$ exists and is the identify function. From Proposition A.1, the optimal contract can therefore be characterized as

$$\frac{1}{U'(s(x))} = \lambda + \frac{x-s(x)-\eta}{U(s(x))-\nu}. \tag{26}$$

In summary, there is a large class of additively separable cost functions for which the use of likelihood ratios is completely negated by the nature of the conflict of interest between the principal and the agent. We leave it to future research to examine the interaction between the conflict of interest component and the likelihood ratio component when $\tilde{v}$ does not exist, and to more sharply identify the properties of $v$ that allow for the existence of $\tilde{v}$.

### A.2 Multiplicative separability

We now show that our results are robust to relaxing additive separability. Let the agent's cost function be given by some function $V(f)$. The following proposition provides sufficient conditions on $V(f)$ such that then the main insight from Proposition 2 holds – the optimal contract is conditioned on $x$ directly rather than through the likelihood ratio.

**Proposition A.2** *Let the agent choose the distribution over $x$ at personal cost $V(f)$. Assume that $V(f)$ is convex in $f$, twice differentiable with respect to $f$, and admits the existence of a function $\tilde{V}$ such that $\tilde{V}\left(\frac{\partial V(f)}{\partial f(x)}, \frac{\partial V(f)}{\partial f(\tilde{x})}\right) = f(x)\frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)}$. Then the following expression characterizes the optimal contract solving the principal's program.*

$$\frac{\int\left(\frac{1}{U'(s(\tilde{x}))}-\lambda\right)\tilde{V}(U(s(x))-\nu,U(s(\tilde{x}))-\nu)d\tilde{x}}{\int\tilde{V}(U(s(x))-\nu,U(s(\tilde{x}))-\nu)d\tilde{x}} = \lambda + \frac{x-s(x)-\eta}{\int\tilde{V}(U(s(x))-\nu,U(s(\tilde{x}))-\nu)d\tilde{x}} \tag{27}$$

The contract does not depend on $f$ and is instead some transformation of the principal's objective. The following corollary identifies a large class of multiplicatively separable cost functions satisfying the condition specified in Proposition A.2.

**Corollary A.1** *Let the principal's objective, $x$, be observable, and let the agent control $f(x)$ with a personal cost given by the geometric product integral $V(f) = \prod v_x(f(x))^{dx} = e^{\int \ln(v_x(f(x)))dx}$. Then if for every $x$ the function $v_x(f(x))$ is homogeneous of any degree, there exists a $\tilde{V}$ as specified in Proposition A.2 and the optimal contract is characterized by equation (27).*

# B Proofs

*Proof of Lemma 1.* It follows immediately from (5) that for a given contract $s$, the incentive compatible $f$ is unique. To show the converse, suppose the principal wants to implement a particular distribution $f$ satisfying $f(x) > 0$ and $\int f(x) = 1$ for all $x$ by designing a contract $s$. To show there exists a contract that will implement this $f$, consider the contract $s(x) = U^{-1}\left(\ln\left(\frac{f(x)}{g(x)}\right) + 1\right)$. Given this contract, the agent chooses $\tilde{f}$ to maximize

$$\begin{aligned} \max_{\tilde{f}} \quad & \int U(s(x))\tilde{f}(x)dx - \int \ln\left(\frac{\tilde{f}(x)}{g(x)}\right)\tilde{f}(x)dx \\ \text{s.t.} \quad & 1 = \int \tilde{f}(x)dx. \end{aligned} \tag{28}$$

Suppose that the constraint does not bind. Pointwise optimization of the Lagrangian yields the following characterization of the optimal action, $\tilde{f}$.

$$\tilde{f}(x) = g(x)e^{U(s(x))-1} = g(x)e^{U(U^{-1}(\ln\left(\frac{f(x)}{g(x)}\right))+1)-1} = g(x)e^{\ln\left(\frac{f(x)}{g(x)}\right)} = g(x)\frac{f(x)}{g(x)} = f(x).$$

Because $\int \tilde{f}(x)dx = \int f(x)dx = 1$, this solution satisfies the constraint. Thus the contract $s(x) = U^{-1}\left(\ln\left(\frac{f(x)}{g(x)}\right) + 1\right)$ implements $f$.

To show uniqueness up to addition by a constant, assume that there is some alternative contract $\tilde{s}(x)$ that implements $f$. Letting $\nu$ denote the Lagrange multiplier on the constraint, the agent's first-order condition is given by

$$U(\tilde{s}(x)) = \ln\left(\frac{\tilde{f}(x)}{g(x)}\right) + 1 + \nu = \ln\left(\frac{f(x)}{g(x)}\right) + 1 + \nu = U(s(x)) + 1 + \nu, \tag{29}$$

where the second equality follows by the assumption that $\tilde{s}$ implements $f$ and the third follows from the definition $s(x) = U^{-1}\left(\ln\left(\frac{f(x)}{g(x)}\right) + 1\right)$. That is, $U(\tilde{s}(x))$ is equal to $U(s(x))$ plus a constant.

□

*Proof of Proposition 1.* Setting $U(s(x)) = s(x)$ in equation (5) yields the risk-neutral agent's incentive compatible action as

$$f(x) \quad = \quad g(x)e^{s(x)-\nu-1}. \tag{30}$$

Then the principal's program is as follows.

$$\begin{array}{rl} \max\limits_{s,f} & \int xf(x)dx - \int s(x)f(x)dx \\ \text{s.t.} & \nu + \int (s(x)-\nu)f(x)dx - \int f(x)\ln\left(\frac{f(x)}{g(x)}\right)dx \geq \bar{H} \\ & f(x) = g(x)e^{s(x)-\nu-1} \text{ for all } x \\ & 1 = \int f(x)dx \end{array} \tag{31}$$

Substitute the IC constraint into the objective function and the other constraints. Then substituting the final constraint into the IR constraint yields $\nu+1 \geq \bar{H}$, allowing the principal's program to be rewritten as follows.

$$\begin{array}{rl} \max\limits_{s} & \int xg(x)e^{s(x)-\bar{H}}dx - \int s(x)g(x)e^{s(x)-\bar{H}}dx \\ \text{s.t.} & 1 = \int g(x)e^{s(x)-\bar{H}}dx \end{array} \tag{32}$$

Letting $\eta$ be the multiplier on the constraint, pointwise optimization with respect to $s$ yields (8). Substituting (8) and $\nu = \bar{H}-1$ into (30) gives (9). Finally, substituting (9) into the constraint $1 = \int f(x)dx$ gives $\eta = \ln\left(\int g(x)e^{x-\bar{H}-1}dx\right)$.

□

*Proof of Proposition 2.* Pointwise optimization of (7) with respect to $f$ at $x$ allows us to solve for $\mu(x)$ in closed form as $\mu(x) = f(x)\,(x - s(x) - \eta)$. Substituting this into (10) yields equation (11). To show that $s(x)$ is increasing in $x$, we first rewrite equation (11) as

$$\frac{1}{U'(x-r(x))} \quad = \quad \lambda - \eta + r(x), \tag{33}$$

where $r(x) \equiv x - s(x)$ is the principal's residual. By our assumption that $U''(\cdot) < 0$, $U'(x - r(x))$ is increasing in $r(x)$ for a fixed $x$, so $\frac{1}{U'(x-r(x))}$ is decreasing in $r(x)$ for a fixed $x$. Now hold $r(x)$ fixed and consider an increase in $x$. This increases the left-hand side while the right-hand side is constant. Then to maintain (33), a decrease in

$r(x)$ must accompany an increase in $x$. Then $s(x)$ is increasing in $x$.

□

*Proof of Proposition 3.* Given some contract $s(x, \mathbf{y})$, the agent chooses $f(x, \mathbf{y})$ to maximize his expected utility from compensation less his personal cost. The agent's program is given by

$$\begin{aligned} \max_f \quad & \int U(s(x,\mathbf{y}))f(x,\mathbf{y})d(x,\mathbf{y}) - \int \ln\left(\tfrac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) f(x,\mathbf{y})d(x,\mathbf{y}) \\ \text{s.t.} \quad & 1 = \int f(x,\mathbf{y})d(x,\mathbf{y}). \end{aligned} \tag{34}$$

Pointwise optimization yields the following incentive compatible action:

$$\begin{aligned} & U(s(x,\mathbf{y})) = \ln\left(\tfrac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) + 1 + \nu \\ \iff & f(x,\mathbf{y}) = g(x,\mathbf{y})e^{U(s(x,\mathbf{y}))-1-\nu} \end{aligned} \tag{35}$$

This unique incentive compatible action satisfies the following.

$$\begin{aligned} f(x,\mathbf{y}) = g(x,\mathbf{y})e^{U(s(x,\mathbf{y}))-\nu-1} \quad & \implies \quad f(x) = \int f(x,\tilde{\mathbf{y}})d\tilde{\mathbf{y}} = \int g(x,\tilde{\mathbf{y}})e^{U(s(x,\tilde{\mathbf{y}}))-\nu-1}d\tilde{\mathbf{y}} \\ & \implies \quad f(\mathbf{y}|x) = \tfrac{f(x,\mathbf{y})}{f(x)} = \tfrac{g(x,\mathbf{y})e^{U(s(x,\mathbf{y}))-\nu-1}}{\int g(x,\tilde{\mathbf{y}})e^{U(s(x,\tilde{\mathbf{y}}))-\nu-1}d\tilde{\mathbf{y}}}. \end{aligned}$$

It follows that $f(\mathbf{y}|x) = g(\mathbf{y}|x)$ if and only if $e^{U(s(x,\mathbf{y}))-\nu} = \frac{1}{g(x)} \int g(x,\tilde{\mathbf{y}})e^{U(s(x,\tilde{\mathbf{y}}))-\nu-1}d\tilde{\mathbf{y}}$. Because the right hand side is constant in $\mathbf{y}$, this expression holds for all $\mathbf{y}$ if and only if $s(x,\mathbf{y}) = s(x)$. □

*Proof of Proposition 4.* Given the IC action (35), the principal's program is as follows.

$$\begin{aligned} \max_{s,f} \quad & \int (x - s(x,\mathbf{y}))f(x,\mathbf{y})d(x,\mathbf{y}) \\ \text{s.t.} \quad & \nu + \int U(s(x,\mathbf{y}) - \nu)f(x,\mathbf{y})d(x,\mathbf{y}) - \int \ln\left(\tfrac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) f(x,\mathbf{y})d(x,\mathbf{y}) \\ & U(s(x,\mathbf{y})) = \ln\left(\tfrac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) + 1 + \nu \text{ for all } (x,\mathbf{y}) \\ & 1 = \int f(x,\mathbf{y})d(x,\mathbf{y}) \end{aligned} \tag{36}$$

Let $\lambda$, $\mu(x,\mathbf{y})$, and $\eta$ denote the Lagrange multipliers on the constraints. Pointwise optimization with respect to $s$ yields:

$$\frac{1}{U'(s(x,\mathbf{y}))} = \lambda + \mu(x,\mathbf{y})\frac{1}{f(x,\mathbf{y})}. \tag{37}$$

Pointwise optimization with respect to $f$ gives $\mu(x,\mathbf{y}) = f(x,\mathbf{y})(x - s(x,\mathbf{y}) - \eta)$. Substituting this into (37) yields (15). To see that (15) can be maintained for all $\mathbf{y}$ only if $s(x,\mathbf{y}) = s(x)$ for all $\mathbf{y}$, notice first that because $U''(\cdot) < 0$, $U'(s(x,\mathbf{y}))$ is weakly decreasing in $s(x,\mathbf{y})$. Then $\frac{1}{U'(s(x,\mathbf{y}))}$ is increasing in $s(x,\mathbf{y})$ for a fixed $(x,\mathbf{y})$. Since the right-hand side of (15) is decreasing in $s(x,\mathbf{y})$, the left and right hand side move in opposite directions in response to variation in $s(x,\mathbf{y})$; therefore, because $\mathbf{y}$ enters the expression only through $s(x,\mathbf{y})$ the equation can not be maintained if $s(x,\mathbf{y})$ is allowed to vary in $\mathbf{y}$. Having established that $s(x,\mathbf{y}) = s(x)$ for all $\mathbf{y}$, it follows directly from the proof of Proposition 2 that $s$ is increasing in $x$.

□

*Proof of Lemma 3.* Given a contract $s(\mathbf{y})$, the agent's program is as follows.

$$\begin{aligned} \max_f \quad & \int U(s(\mathbf{y}))f(x,\mathbf{y})d(x,\mathbf{y}) - \int \ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) f(x,\mathbf{y})d(x,\mathbf{y}) \\ \text{s.t.} \quad & 1 = \int f(x,\mathbf{y})d(x,\mathbf{y}) \end{aligned} \tag{38}$$

Pointwise optimization yields the following incentive compatible action:

$$f(x,\mathbf{y}) = g(x,\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu} \tag{39}$$

It follows that the incentive compatible $f$ given $s$ is unique. Because both $g$ and the exponential function are positive, $f(x,\mathbf{y}) > 0$. Moreover,

$$f(\mathbf{y}) = \int f(x,\mathbf{y})dx = \int g(x,\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}dx = g(\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}, \tag{40}$$

which implies that $f(x|\mathbf{y}) = \frac{f(x,\mathbf{y})}{f(\mathbf{y})} = \frac{g(x,\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}}{g(\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}} = \frac{g(x,\mathbf{y})}{g(\mathbf{y})} = g(x|\mathbf{y})$.

That is, any contract written only on $\mathbf{y}$ motivates a unique action satisfying $f(x|\mathbf{y}) = g(x|\mathbf{y})$.

To show the converse, suppose the principal wants to implement a particular $f$ satisfying $f(x,\mathbf{y}) > 0$, $f(x|\mathbf{y}) = g(x|\mathbf{y})$, and $\int f(x,\mathbf{y}) = 1$ for all $(x,\mathbf{y})$. To show there exists a contract that will implement this $f$, consider the contract $s(\mathbf{y}) = U^{-1}\left(\ln\left(\frac{f(\mathbf{y})}{g(\mathbf{y})}\right)+1\right)$. Given this contract, the agent chooses $\tilde{f}$ to maximize

$$\begin{aligned} \max_{\tilde{f}} \quad & \int U(s(\mathbf{y}))\tilde{f}(x,\mathbf{y})d(x,\mathbf{y}) - \int \ln\left(\frac{\tilde{f}(x,\mathbf{y})}{g(x,\mathbf{y})}\right)\tilde{f}(x,\mathbf{y})d(x,\mathbf{y}) \\ \text{s.t.} \quad & 1 = \int \tilde{f}(x,\mathbf{y})d(x,\mathbf{y}). \end{aligned} \tag{41}$$

Suppose that the constraint does not bind. Pointwise optimization yields

$$\begin{aligned} \tilde{f}(x,\mathbf{y}) &= g(x,\mathbf{y})e^{U(s(\mathbf{y}))-1} \\ &= g(x|\mathbf{y})g(\mathbf{y})e^{U\left(U^{-1}\left(\ln\left(\frac{f(\mathbf{y})}{g(\mathbf{y})}\right)\right)+1\right)-1} \;=\; f(x|\mathbf{y})g(\mathbf{y})\frac{f(\mathbf{y})}{g(\mathbf{y})} \;=\; f(x,\mathbf{y}), \end{aligned}$$

where the second equality holds by substituting in $s(\mathbf{y})$ and the third holds because $f(x|\mathbf{y}) = g(x|\mathbf{y})$. Because $\int \tilde{f}(x,\mathbf{y})d(x,\mathbf{y}) = \int f(x,\mathbf{y})d(x,\mathbf{y}) = 1$, this solution satisfies the constraint. Thus the contract $s(\mathbf{y}) = U^{-1}\left(\ln\left(\frac{f(\mathbf{y})}{g(\mathbf{y})}\right)+1\right)$ implements $f$.

To show uniqueness up to addition by a constant, assume that there is some alternative contract $\tilde{s}(x)$ that implements $f$. The agent's first-order condition is

$$U(\tilde{s}(\mathbf{y})) = \ln\left(\frac{\tilde{f}(x,\mathbf{y})}{g(x,\mathbf{y})}\right) + 1 + \nu = \ln\left(\frac{f(\mathbf{y})}{g(\mathbf{y})}\right) + 1 + \nu = U(s(\mathbf{y})) + 1 + \nu, \tag{42}$$

where the second equality follows by the assumption that $\tilde{s}$ implements $f$ with $f(x|\mathbf{y}) = g(x|\mathbf{y})$ and the third equality follows from the definition of $s(\mathbf{y})$. That is, $U(\tilde{s}(\mathbf{y}))$ is equal to $U(s(\mathbf{y}))$ plus a constant. □

*Proof of Proposition 5.* Given IC action (39), the principal's program is as follows.

$$\begin{aligned} \max_{s,f} \quad & \int (x - s(\mathbf{y}))f(x,\mathbf{y})d(x,\mathbf{y}) \\ \text{s.t.} \quad & \nu + \int \left(U(s(\mathbf{y})) - \nu\right) f(x,\mathbf{y})d(x,\mathbf{y}) - \int \ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) f(x,\mathbf{y})d(x,\mathbf{y}) \geq \bar{H} \\ & U(s(\mathbf{y})) = \ln\left(\frac{f(x,\mathbf{y})}{g(x,\mathbf{y})}\right) + 1 + \nu \text{ for all } (x,\mathbf{y}) \\ & 1 = \int f(x,\mathbf{y})d(x,\mathbf{y}) \end{aligned} \tag{43}$$

Pointwise optimization of (43) over $s(\mathbf{y})$ yields

$$\frac{1}{U'(s(\mathbf{y}))} = \lambda + \int \mu(x,\mathbf{y})dx\frac{1}{f(\mathbf{y})}. \tag{44}$$

Pointwise optimization of (43) over $f(x, \mathbf{y})$ yields $\mu(x, \mathbf{y}) = f(x, \mathbf{y})\,(x - s(\mathbf{y}) - \eta)$. Substituting $\mu(x, \mathbf{y})$ into (44) gives:

$$\frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \int x \frac{f(x,\mathbf{y})}{f(\mathbf{y})} dx - s(\mathbf{y}). \tag{45}$$

Noticing that $\int x \frac{f(x,\mathbf{y})}{f(\mathbf{y})} dx = \int x f(x|\mathbf{y}) dx = E_f[x|\mathbf{y}]$, the above expression reduces to the characterization presented in the proposition.

□

*Proof of Proposition 6.* By the definition of a marginal distribution, we can integrate (39) over $\tilde{x}$ to obtain $f(\mathbf{y}) = g(\mathbf{y})e^{U(s(\mathbf{y}))-\nu-1}$. Then by definition of a conditional distribution, we can divide this into $f(x, \mathbf{y})$ to obtain

$$f(x|\mathbf{y}) \;=\; \frac{g(x,\mathbf{y})e^{U(s(\mathbf{y}))-\nu-1}}{g(\mathbf{y})e^{U(s(\mathbf{y}))-\nu-1}} \;=\; g(x|\mathbf{y}). \tag{46}$$

Then $E_f[x|\mathbf{y}] = \int x f(x|\mathbf{y}) dx = \int x g(x|\mathbf{y}) dx = E_g[x|\mathbf{y}]$, so we can rewrite (17) as

$$\frac{1}{U'(s(\mathbf{y}))} \;=\; \tilde{\lambda} + \mathbb{E}_g[x|\mathbf{y}] - s(\mathbf{y}). \tag{47}$$

Notice that $\mathbf{y}$ enters the contract only through $\mathbb{E}_g[x|\mathbf{y}]$ and $s(\mathbf{y})$. If $\mathbb{E}_g[x|\mathbf{y}]$ varies in $y_i$, then the only way (47) can be maintained for all $\mathbf{y}$ is if $s(\mathbf{y})$ also varies in $y_i$; that is, the contract is conditioned in part on $y_i$. If $\mathbb{E}_g[x|\mathbf{y}]$ does not vary in $y_i$ – that is, if $E_g[x|\mathbf{y}] = E_g[x|\mathbf{y}_{-i}]$ – then (47) is violated for some $y_i$ if $s(\mathbf{y})$ varies in $y_i$.

□

*Proof of Proposition A.1.* If there exists a function $\tilde{v}$ as described, we can rewrite (24) as

$$\frac{1}{U'(s(x))} \;=\; \lambda + \frac{x - s(x) - \eta}{\tilde{v}(v'_x(f(x)))}. \tag{48}$$

Substituting the IC action (21) into (48) gives solution (25). To show that $s(x)$ is increasing in $x$, we first rewrite equation (25) as

$$\frac{1}{U'(x - r(x))} \;=\; \lambda + \frac{r(x) - \eta}{\tilde{v}(U(x - r(x)) - \nu)}, \tag{49}$$

where $r(x) \equiv x - s(x)$ is the principal's residual. By our assumption that $U''(\cdot) < 0$, $U'(x - r(x))$ is increasing in $r(x)$ for a fixed $x$. Then $\frac{1}{U'(x-r(x))}$ is decreasing in $r(x)$ for a fixed $x$. The denominator on the right-hand side of (49) is decreasing in $r(x)$ because $U(x - r(x))$ is decreasing in $r(x)$ and $\tilde{v}(\cdot)$ is everywhere nondecreasing in its argument. Then $\frac{r(x)-\eta}{\tilde{v}(U(x-r(x))-\nu)}$ is increasing in $r(x)$ for a fixed $x$. Now hold $r(x)$ fixed and consider an increase in $x$. This increases the left-hand side of (49) and decreases the right-hand side. Then for (49) to be maintained, an increase in $x$ must be accompanied by a decrease in $r(x)$. Then $s(x)$ is increasing in $x$.

□

*Proof of Proposition A.2.* The agent's program is given as follows.

$$\begin{array}{ll} \max\limits_{f} & \int U(s(x))f(x)dx - V(f) \\ \text{s.t.} & 1 = \int f(x)dx \end{array} \tag{50}$$

Pointwise optimization yields the following expression for the IC action:

$$\frac{\partial V(f)}{\partial f(x)} \;\;=\;\; U(s(x)) - \nu \tag{51}$$

The principal's optimization program is given by the following Lagrangian.

$$\begin{aligned} \mathcal{L}(s, f) \;\; &= \int (x - s(x))f(x)dx \\ &+\lambda\left[\nu + \int (U(s(x)) - \nu)f(x)dx - V(f) - \bar{H}\right] \\ &+ \int \mu(\tilde{x})\left(U(s(\tilde{x})) - \nu - \frac{\partial V(f)}{\partial f(\tilde{x})}\right)d\tilde{x} \\ &+\eta\left[1 - \int f(x)dx\right] \end{aligned} \tag{52}$$

Taking the first-order condition with respect to $f(x)$ yields

$$x - s(x) - \eta - \int \mu(\tilde{x})\frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)}d\tilde{x} = 0 \tag{53}$$

Optimizing (52) pointwise over $s(x)$ yields $\mu(x) = \left(\frac{1}{U'(s(x))} - \lambda\right)f(x)$. Substituting this into (53) and invoking the existence of $\tilde{V}$ gives

$$x - s(x) - \eta - \int \left(\frac{1}{U'(s(\tilde{x}))} - \lambda\right)\tilde{V}\left(\frac{\partial V(f)}{\partial f(x)}, \frac{\partial V(f)}{\partial f(\tilde{x})}\right)d\tilde{x} = 0. \tag{54}$$

Finally, substituting in the agent's first-order condition produces the solution.

□

*Proof of Corollary A.1.* Letting $V(f) = e^{\int \ln(v_x(f(x)))dx}$, the first derivative $\frac{\partial V(f)}{\partial f(x)}$ and the cross partial $\frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)}$ can be obtained as follows.

$$\begin{aligned} \frac{\partial V(f)}{\partial f(\tilde{x})} &= \frac{v'_{\tilde{x}}(f(\tilde{x}))}{v_{\tilde{x}}(f(\tilde{x}))} V(f) \\ \frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)} &= \left(\frac{v'_x(f(x))}{v_x(f(x))}\right) \frac{\partial V(f)}{\partial f(\tilde{x})} \end{aligned} \tag{55}$$

Multiplying the cross partial by $f(x)$ gives

$$f(x) \cdot \frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)} = \left(\frac{f(x) v'_x(f(x))}{v_x(f(x))}\right) \frac{\partial V(f)}{\partial f(\tilde{x})} \tag{56}$$

Suppose that for all $x$, $v_x(f(x))$ is homogeneous of degree $k$; that is, $v_x(\alpha f(x)) = \alpha^k v_x(f(x))$ where $\alpha$ is a constant. Then by Euler's theorem, $f(x)v'_x(f(x)) = k v_x(f(x))$, and therefore the term $\frac{f(x) \cdot v'_x(f(x))}{v_x(f(x))}$ in (56) is equal to the constant $k$. Then $f(x) \cdot \frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)} = k \frac{\partial V(f)}{\partial f(\tilde{x})}$, and thus there exists some function $\tilde{V}$ such that $\tilde{V}\left(\frac{\partial V(f)}{\partial f(x)}, \frac{\partial V(f)}{\partial f(\tilde{x})}\right) = f(x) \frac{\partial^2 V(f)}{\partial f(\tilde{x})\partial f(x)}$. It follows from Proposition A.2 that the solution is given by (27).

□