Contracting on Aggregated Accounting Estimates

Jonathan Bonham
University of Chicago
Booth School of Business

Amoray Riggs-Cragun
University of Chicago
Booth School of Business

September 23, 2021

Abstract

Using a principal-agent framework in which the agent chooses the joint distribution over all contractible and non-contractible signals, we provide a theoretical justification for contracting on aggregated accounting estimates. The optimal contracting process can be decomposed into three stages: estimating individual items that the principal values, aggregating those estimates using the weights in the principal's objective (as opposed to weights driven by sensitivity or precision), and compensating the agent based on the aggregated estimate. In a highly tractable specification of our model in which normal distributions arise endogenously, we show that optimal measurement rules are conservative yet produce unbiased estimates and we rationalize the immediate expensing of R&D, the capitalization of PP&E, the accrual of credit sales, and fair value accounting for certain financial investments.

Keywords: Optimal contracting, aggregation, accounting measurement.

We appreciate helpful comments from Jeremy Bertomeu, Robert Göx, Christian Leuz, Valeri Nikolaev, Haresh Sapra, Abbie Smith, and participants in the DAR & DART Accounting Theory Seminar, the Chicago Booth Accounting Research Workshop, and the Accounting and Economics Society webinar. 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

Executive compensation contracts routinely tie CEO pay to highly aggregated accounting measures such as total revenue or net income.[1] However, classical agency theory suggests that aggregated accounting estimates are inefficient contracting measures, because these aggregates are constructed by simply adding up account values, not by weighting the accounts according to their relative sensitivities and precision as in Banker and Datar [1989]. Absent imposing additional frictions to the moral hazard problem, agency theory has yet to explain why aggregated accounting measures like earnings are so widely used in executive compensation contracts. Using an unconventional but well-grounded approach to moral hazard where the agent can affect the entire shape of the outcome distribution, we show that contracting on aggregated accounting estimates can be just as efficient as contracting on all of the information used to measure their underlying components.

The atypical ingredient in our model is in how the agent's action is represented. Beginning with the seminal work of Holmström [1979], agency theory has been preponderated by the *parameterized distribution formulation* of the moral hazard problem, which models the agent's actions as affecting the parameter(s) of a distribution whose form is outside of the agent's control. For example, it is common to assume that the agent chooses the first moment of a normal distribution (e.g. Holmström and Milgrom [1991], Feltham and Xie [1994]); here, the agent can mean-shift the normal distribution but has no control over its shape. Less commonly, some papers assume that the agent influences both the first and second moments of a normal distribution (e.g. Meth [1996]), but even here the agent's influence is quite limited. He has no way of affecting skewness or kurtosis or of introducing discontinuities; by assumption, the *only* possibility is a normal distribution. The same limitations apply when the parametric

[1] For example, De Angelis and Grinstein [2015] report that four in five CEO performance-based awards are tied to accounting metrics (Figure 1), with income and sales measures among the most common.

form is unspecified, for instance when the agent is modeled as choosing generic parameter a (often called "effort") in the distribution over x, denoted $f(x; a)$. The agent's choice of a entails choosing a distribution from a restricted parametric set, and there is nothing he can do to break free of the functional form of $f(x; a)$.

We see the parametric approach as quite limiting, and in this paper we take an alternative (though not novel) approach. We use what Hart and Holmström [1987] termed the *generalized distribution formulation* of the moral hazard problem, which assumes that the agent can implement *any* distribution nonparametrically. This generalized approach, in our view, better reflects the flexibility that firm managers have in running a company. The sheer number of books proffering business advice (over 80,000 titles on Amazon) suggests that a CEO's inputs are far more complex than simply exerting effort along one or more dimensions (else those books could be replaced by the maxim, "Try hard.") Many empirical papers in the accounting literature take it as given that executives can influence distributions nonparametrically; for example, discontinuities around zero or analyst forecasts are often attributed to the management of real or measured earnings (e.g., Burgstahler and Dichev [1997], Roychowdhury [2006]).

We use the generalized distribution approach to study two central accounting issues: *measurement* and *aggregation*. *Measurement* is the process by which the principal estimates unobservable variables, and *aggregation* is the process by which multiple measures are combined into a composite one. In particular, we study estimated or aggregated performance measures that are *efficient* or (weakly) optimal for contracting purposes, meaning that there is no loss to contracting on an estimate or aggregate relative to contracting on all of the underlying information used to construct it.

The standing authority on optimal aggregation is Banker and Datar [1989]. In their seminal paper, Banker and Datar use the parametric approach to show that optimal linear aggregates are formed by weighting the underlying measures according to their relative *precision* and *sensitivity*. That is, measures that have lower variance or are more responsive to managerial actions should receive relatively higher weight. Lambert [2001] observes that this result suggests that accounting aggregates, which are *equally*-weighted sums or differences of underlying accounts, are unlikely to be efficient for contracting. Income statement line items such as sales, cost of goods sold and depreciation are likely to vary widely in their variance and in their sensitivities to CEO effort. Because income statement aggregates such as earnings before taxes (EBT) or net income subtract expenses from revenues with no regard for differences in the sensitivity or precision of these components, Banker and Datar's result implies that these accounting aggregates are inefficient for contracting.

We revisit optimal aggregation using the generalized distribution approach. We show that if the agent controls $f(\mathbf{x})$, and if the principal's *objective* (the thing she wishes to maximize) is linear in $\mathbf{x}$, then the optimal linear aggregate is given simply by the principal's objective. That is, the optimal weight on each variable is determined solely by how strongly the principal values it. For example, if shareholders care equally about each revenue and expense item, then an optimal compensation contract can be written on net income, which is an equally-weighted aggregate of all revenues and expenses. Alternatively, if shareholders care more about revenues than they do about line items further down the income statement, then the optimal compensation contract will weight revenues more heavily. These results hold regardless of how noisy the various revenue and expense items are or how sensitive they are to the manager's actions.

Our aggregation results may be of particular interest to empirical compensation research, which often uses the sensitivity-precision result from Banker and Datar [1989] to form predictions about executive compensation contracts. As noted by Bushman and Smith [2001], it is difficult to empirically operationalize precision and (especially) sensitivity, and consequently, the findings from this literature are mixed and sensitive to empirical specification. For example, Core, Guay, and Verrecchia [2003] find that the relative weights on performance measures are decreasing in relative variance when looking at CEO cash compensation but *increasing* in relative variance when considering CEO total compensation, suggesting that "existing findings on cash pay cannot be interpreted as evidence supporting standard agency predictions."

Before signals can be aggregated, they must be measured. The underlying firm fundamentals that shareholders care about – such as true revenue, true income, or investment value – are unobservable constructs that are estimated by the accounting system. To capture this reality, we expand our model to a setting where the principal cannot observe the variables she intrinsically values, $\mathbf{x}$, but she can observe some other set of variables, $\mathbf{y}$, such as the timing and amounts of individual transactions. The agent influences both sets of variables by choosing $f(\mathbf{x}, \mathbf{y})$. We show that the principal's solution can be broken into three stages: measurement, aggregation, and contracting. In the measurement stage, the principal uses information in $\mathbf{y}$ to construct an *unbiased* estimate of each element in $\mathbf{x}$; that is, the estimates are representationally faithful. In the aggregation stage, as before, the principal aggregates the estimates according to their corresponding weights in her objective. For example, if she cares equally about the revenues from each product, these revenues are simply summed up into total revenues. Finally, in the contracting stage, the principal conditions the agent's compensation contract on one or more of the aggregated estimates. This three-step decomposition rationalizes contracting on aggregated accounting measures in practice: accounting systems estimate things that shareholders care about, then sum these estimates into aggregate measures, and executive compensation contracts are written on those aggregates.

Accounting standard setters are particularly concerned with the *measurement* stage, where the observable signals $\mathbf{y}$ are used to produce unbiased estimates of underlying fundamentals $\mathbf{x}$. To explore the measurement process, we develop a highly tractable specification of our model in which the agent's equilibrium action is to mean-shift a normal distribution. (We do not impose normality on the solution; it arises endogenously despite the agent's ability to implement any distribution imaginable.) Using this specification, we show that optimal measurement rules are conservative, whereas the resulting measures themselves are unbiased. Because the manager can game the transaction characteristics $\mathbf{y}$ in an attempt to overstate underlying performance, the optimal measurement rule understates $\mathbf{y}$ to arrive at an unbiased estimate of $\mathbf{x}$. Consistent with the findings of Gao [2013] and the intuition suggested by Watts [2003], conservative accounting offsets managerial manipulation to produce unbiased performance measures.

One important measurement issue in accounting is how to measure investments with uncertain returns. We present a simple application of our model which suggests that optimal measurement depends on the availability of reliable evidence about future returns. In particular, if reliable evidence is difficult to produce, the optimal measurement rule for a given investment depends on the likelihood that the investment will produce future returns. When the correlation between an investment and future returns is low, the optimal measurement rule reports the cash outlay and makes no attempt to estimate future returns, consistent with the immediate expensing of highly uncertain investments such as R&D. When the correlation between investment and future return is high, net income optimally includes some unrealized income, consistent with the accrual of revenues after inventory has been delivered but before cash is collected. When the correlation between investment and future return is moderate, the optimal measurement rule deducts some, but not all, of the investment outlay from net income. With the caveat that our highly stylized model considers only a single period, this result resembles the capitalization and depreciation of PP&E investments, which arguably are more strongly correlated with future returns than R&D but less so than sales of inventory on credit. Finally, if reliable evidence about future returns is easy to produce, the optimal measurement rule reports investments at fair value, *regardless* of the correlation between the investment and its future return. For example, if the present value of a financial investment can be reliably estimated, it is reported at fair value regardless of whether it is likely to generate future returns.

The paper proceeds as follows. Section 2 discusses related work. In section 3, we revisit the question of optimal aggregation studied by Banker and Datar [1989]. Using the generalized distribution approach, we show that the efficient linear aggregator weights the signals according to their weights in the principal's objective, not according to their relative sensitivities and precision. In section 4, we extend the model to a setting that seems more descriptive of the accounting process: the principal cares about a set of *unobservable* firm fundamentals, $\mathbf{x}$, and can observe a large set of other signals (e.g. transaction characteristics), $\mathbf{y}$. We show that contracting on aggregated estimates of $\mathbf{x}$ is just as efficient as contracting on $\mathbf{y}$, providing a theoretic justification for contracting on aggregated accounting estimates like net income. In section 5, we use a tractable specification of our model to investigate how the estimates of $\mathbf{x}$ are formed. We provide stylized applications that rationalize conservative measurement rules (section 5.1) as well as immediate expensing of R&D, capitalization of fixed assets, accrual of credit sale revenue, and fair value accounting (section 5.2). Section 6 concludes and provides suggestions for future empirical research. Proofs are in the appendix.

2 Related Work

Our paper is closest in topic to Banker and Datar [1989], who also study optimal aggregation in a moral hazard setting. In particular, we are interested in the following question: Given a large number of signals under an agent's control, is there a way to linearly aggregate the signals such that there is no loss to conditioning the contract on the aggregate relative to conditioning the contract on all of the underlying signals? Some papers in accounting introduce frictions that make aggregation strictly optimal. For example, Amershi and Cheng [1989] derive a demand for aggregation by assuming it is costly to design and implement contracts based on disaggregated data. Aggregation is optimal when the cost of contracting on many variables is outweighed by the cost of the information loss that aggregation entails. More recently, Arya and Glover [2014] suggest several settings in which the information loss from aggregation is actually beneficial. Our paper and Banker and Datar [1989] contain only the standard moral hazard frictions – risk aversion and unobservability of the agent's actions – and study how *weakly* optimal aggregates are formed. Trivially, these aggregates would strictly dominate if we modeled some cost to contracting on disaggregated information, such as a "cost of complexity" that increases in the number of measures included in the contract.

Extending the setting studied by Banker and Datar [1989], we show that if the variables the principal cares about are unobservable, the optimal contract can be broken into three stages – measurement, aggregation and contracting. In the measurement stage, an *efficient measurement rule* uses observable data (i.e. transactions or economic events) to produce a weakly optimal estimate, such that there is no loss to contracting on the estimate relative to contracting on the underlying data. Our decomposition of the solution into separate measurement and contracting stages is related to Leuz [1999], who points out that contingencies embedded in the accounting function could instead be included directly in contracts. We show only weak optimality of separating the measurement and compensation functions, but as Leuz [1999] suggests, one could introduce contracting costs to make the separation strictly optimal. For example, if the principal contracts with multiple parties and there are costs to designing and implementing each contract, there may be "returns to scale" from using the same measures in multiple contracts.

Much of the accounting literature models measurement as a biased, one-step mapping from fundamentals to performance. By contrast, our model follows a *two-step approach* as proposed by Gao [2013]. The agent chooses $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})f(\mathbf{y}|\mathbf{x})$, and fundamentals $\mathbf{x}$ are distributed according to the marginal distribution $f(\mathbf{x})$. These fundamentals are linked to accounting measures in two steps. Fundamentals $\mathbf{x}$ map to transaction characteristics $\mathbf{y}$ according to the conditional distribution $f(\mathbf{y}|\mathbf{x})$, and transaction characteristics are mapped to performance measures according to an *unbiased* estimation process, $\hat{\mathbf{x}} \equiv \mathbb{E}[\mathbf{x}|\mathbf{y}]$, determined as part of the principal's solution.

The key difference between our paper and conventional agency theory papers like Banker and Datar [1989] is in how the agent's action is modeled. While the bulk of the literature uses the *parameterized distribution formulation*, we use the *generalized distribution formulation.*

These terms were coined by Hart and Holmström [1987] in their enlightening review of the early agency literature, where they define and defend the generalized approach as follows (Hart and Holmström [1987], pp.78-79).

> Since the agent [in the parametric approach] 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. This way of looking at the principal's problem is also very general. It covers situations where the agent may observe some information about the cost of his actions, or the expected returns from his actions, before actually deciding what to do; in other words, cases of hidden information. To see this, simply note that whatever strategy the agent uses for choosing actions contingent on information he observes, the strategy will in reduced form map into a distribution choice... Thus, ex ante strategic choices are equivalent to distribution choices in some [probability simplex] P.

The first paper to use the generalized approach was Holmström and Milgrom [1987] (though ironically, that paper is better known as the birthplace of "LEN," a set of modeling assumptions that falls squarely under the parametric approach). In addition to the hidden information example described in the quote above, Holmström and Milgrom [1987] justify the generalized approach with an example in which the agent acts continuously throughout the period, conditioning his action on a privately observed continuous state variable; they argue (and Hébert [2018] shows formally) that this setting can be represented in reduced form as the agent choosing an unconditional distribution at the outset.

Since Holmström and Milgrom [1987], very few papers have used the generalized distribution formulation. An important recent exception is Hébert [2018], who studies optimal security design. He pairs the generalized approach with a novel cost function in which the agent's disutility from implementing some distribution f depends on the divergence of f from a cost-minimizing reference distribution g, and he provides a micro-foundation for this pairing. We adopt Hébert's intuitive cost function in this paper and in Bonham and Riggs-Cragun [2021], a close predecessor to this paper.

In Bonham and Riggs-Cragun [2021], we show that the generalized approach produces optimal contracts that do not depend on likelihood ratios. The agent's pay is conditioned *directly* on the principal's objective, rather than on what the realized outcome says about the agent's action. That intuition holds throughout the present paper as well; likelihood ratios do not appear in our solutions. We extend Bonham and Riggs-Cragun [2021] to an accounting-oriented setting in two ways. First, while in Bonham and Riggs-Cragun [2021] we assume that the principal cares about a single variable, x, here we assume that the principal cares about *many* variables, $\mathbf{x}$. This allows us to study optimal aggregation, an important feature of accounting. Second, in this paper we develop a specification of the model that produces closed-form solutions, uniquely optimal linear contracts, and normal distributions. This tractability allows us to investigate optimal measurement, another issue central to accounting.

In addition to the papers already mentioned, we know of a few others that use 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; and Hemmer [2017] uses a binary version of it to study relative performance evaluation. Finally, Bonham [2020] uses the generalized approach to study how measurement and contracts shape productive incentives. In Bonham [2020], the agent has distributional control only over the principal's objective, x, which is assumed to be non-contractible. The relationship between the objective, x, and the contractible signal, y, is beyond the agent's control. By contrast, in Bonham and Riggs-Cragun [2021] and in the present paper we assume that the agent can influence the joint distribution over both the objective $\mathbf{x}$ and observable signals $\mathbf{y}$, allowing us to take a two-step approach to modeling measurement and to study issues like window dressing.

3 Optimal Weighting of Performance Measures

A principal cares about a set of variables, $\mathbf{x} \equiv (x_1 \ldots x_m)^T$. She might value some variables differently than others; we say that the principal values $\mathbf{x} \in \mathbb{R}^m$ according to her *objective*, $B(\mathbf{x})$, where $B : \mathbb{R}^m \to \mathbb{R}$. The principal hires an agent to take unobservable actions that stochastically improve $\mathbf{x}$. To induce the agent to act in the principal's interest, she designs a contract, $s(\mathbf{x})$, where $s : \mathbb{R}^m \to \mathbb{R}$. Let $U(s)$ be the (weakly) risk averse agent's utility from compensation, where $U : \mathbb{R} \to \mathbb{R}$ satisfies $U'(\cdot) > 0$ and $U''(\cdot) \leq 0$. Denote the agent's utility from outside options by $\bar{U}$. Assume that the principal is risk neutral such that her utility is equal to her net payoff, $B(\mathbf{x}) - s(\mathbf{x})$.

We study two questions in this section. First, under what conditions can the optimal contract be conditioned on a linear aggregate of the performance measures $x_1, \ldots, x_m$ with no loss to conditioning the contract on every individual measure? Second, in cases where the contract can be conditioned on such an aggregate, what are the optimal linear weights on each measure? We first present the classic result from Banker and Datar [1989] under the parametric approach and then revisit the problem by invoking the generalized distribution approach.

3.1 The classic parameterized distribution approach

Under the classic approach taken by Banker and Datar [1989], the agent's action serves as a parameter in the joint distribution over $\mathbf{x}$. Let the agent choose $a \in A \subseteq \mathbb{R}$, and let a parameterize the joint probability density function $f(\mathbf{x}; a)$. Exerting effort imposes a personal cost on the agent of $V(a)$, where the function $V : \mathbb{R} \to \mathbb{R}$ is increasing convex. Assume that the agent's utility is additively separable in his compensation and personal cost, so that we can write his total utility as $U(s) - V(a)$.

The principal's goal is to design a contract-action pair, (s, a), that maximizes her net payoff subject to two constraints. First, the agent's expected utility under the proposed scheme must be at least as high as his utility from outside options; this individual rationality (IR) constraint ensures that the agent will participate in the agency. Second, the proposed contract-action pair must be incentive compatible (IC), meaning that the agent will choose the proposed a when faced with the proposed s. The principal's maximization program is given as follows.

$$\begin{aligned} \max_{s,a} \quad & \int (B(\mathbf{x}) - s(\mathbf{x})) f(\mathbf{x}; a) d\mathbf{x} \\ \text{s.t.} \quad & \int \left(U(s(\mathbf{x})) - V(a)\right) f(\mathbf{x}; a) dx \geq \bar{U} \\ \text{and} \quad & \int U(s(\mathbf{x})) f_a(\mathbf{x}; a) d\mathbf{x} = V'(a) \end{aligned} \tag{1}$$

Let λ and μ be the multipliers on the IR and IC constraints, respectively. Standard methods of pointwise optimization yield the following iconic characterization of the optimal sharing rule.

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

The optimal contract is a transformation of $\lambda + \mu \cdot \frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)}$; we can see this more readily by solving for $s(\mathbf{x})$.

$$s(\mathbf{x}) = U'^{-1}\left(1/\left(\lambda + \mu \cdot \frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)}\right)\right). \tag{3}$$

Banker and Datar [1989] point out that this solution can be decomposed into two stages: (1) aggregating the measures $x_1, \ldots, x_m$ into a single composite measure, $\pi(\mathbf{x})$, and (2) writing a compensation function that is conditioned on that measure.

$$\begin{aligned} \textit{Aggregator} \qquad \pi(\mathbf{x}) & \equiv \lambda + \mu \cdot \frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)} \\ \textit{Contract} \qquad s(\pi(\mathbf{x})) & = U'^{-1}\left(\frac{1}{\pi(\mathbf{x})}\right) \end{aligned} \tag{4}$$

Notice that when the likelihood ratio $\frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)}$ is linear, the optimal aggregator $\pi(\mathbf{x})$ is linear as well.[2] Banker and Datar [1989] define a large class of parametric distributions for which likelihood ratios are linear, including the entire exponential family, providing a rationale for contracting on certain linear aggregates.

Within a subclass of distributions for which some linear aggregator is optimal, Banker and Datar [1989] characterize the relative linear weights on each measure. Let $\pi(\mathbf{x}) = \sum_j \pi_j x_j$ be an optimal linear aggregator. Banker and Datar [1989] show that the weight on measure x_i relative to the weight on measure x_j depends on the relative signal-to-noise ratios of the two measures.

$$\frac{\pi_i}{\pi_j} = \frac{\partial E(x_i|a)/\partial a}{\mathrm{Var}(x_i)} \Big/ \frac{\partial E(x_j|a)/\partial a}{\mathrm{Var}(x_j)} \tag{5}$$

Banker and Datar [1989] call $\partial E(x_i|a)/\partial a$ the *sensitivity* of x_i to the agent's effort and $1/Var(x_i)$ the *precision* of x_i. All else equal, the higher a measure's sensitivity or precision, the larger its relative weight in the optimal aggregator, $\pi(\mathbf{x})$.

The sensitivity-precision result from Banker and Datar [1989] suggests that two performance measures will receive equal weight only if their signal-to-noise ratios are identical. As noted by Lambert [2001], there is something unsatisfying about this from an accounting perspective. Executive compensation contracts are often conditioned on accounting aggregates, which are *equally*-weighted sums and differences of many underlying accounts, and it is unlikely that all of those accounts have identical signal-to-noise ratios. It is therefore unclear from the result why executive compensation contracts are conditioned on aggregated accounting totals rather than on their (readily available and contractible) disaggregated components.

[2]Note that linearity in the aggregator, π, does not imply linearity in the contract, s.

3.2 The generalized distribution approach

We now revisit optimal linear aggregation by invoking the *generalized distribution approach*; that is, we assume that the agent can implement $f(\mathbf{x})$ nonparametrically. This assumption greatly expands the agent's opportunity set relative to the classic parametric approach, giving him as much control over the distribution as the principal has over the contract. Notice that in the classic approach the agent chooses a scalar, $a \in \mathbb{R}$, whereas the principal chooses $s(\mathbf{x})$ for every outcome $\mathbf{x}$. The generalized approach puts the principal and agent on equal footing: the agent chooses the probability of every $\mathbf{x}$ and the principal chooses the pay level for every $\mathbf{x}$.

There are many ways to think about how an agent might "choose a distribution" nonparametrically in practice. In section 2, we referred to two examples provided by Holmström and Milgrom [1987]. Our preferred interpretation is to think of the agent as a CEO who faces an uncountably rich set of opportunities for action. However, it is perhaps more straightforward to simply notice that modeling the agent as implementing $f(\mathbf{x}; a)$ by her choice of a is just a constrained version of modeling the agent as choosing $f(\mathbf{x})$ directly. In the classic model, the agent's choice of a is equivalent to choosing a particular parametric distribution $f(\mathbf{x}; a)$ from the set $\{f(\mathbf{x}; a) | a \in A\}$. The generalized approach assumes that the agent can choose *any* distribution; that is, the agent selects $f(\mathbf{x})$ from $\Delta(\mathbf{x}) \equiv \{f(\mathbf{x}) | \int f(\mathbf{x}) d\mathbf{x} = 1, f(\mathbf{x}) \geq 0 \ \forall \mathbf{x}\}$, where $\Delta(\mathbf{x})$ denotes the space of all probability distributions over $\mathbf{x}$. If one is willing to accept that an agent in the classic approach is capable of implementing a particular $f(\mathbf{x}; a)$, it should not require a radical shift in thinking to accept our assumption that the agent can implement a particular $f(\mathbf{x})$. Of course, some distributions are more difficult to implement than others, which brings us to the agent's cost function.

Assume that the agent implements distribution f at personal cost $V(f)$, a function we redefine here as $V : \Delta(\mathbf{x}) \to \mathbb{R}$. Let $g(\mathbf{x}) \in \Delta(\mathbf{x})$ be the agent's *preferred* or *cost-minimizing* distribution, the one that he would implement if offered zero incentives. Following Hébert [2018], we model the agent's personal cost from implementing f by the *Kullback-Leibler divergence* from g to f, denoted $D(f(\mathbf{x})||g(\mathbf{x}))$.

$$V(f) \;=\; D(f(\mathbf{x})||g(\mathbf{x})) \;\equiv\; \int f(\mathbf{x}) \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x}, \tag{6}$$

where $\int d\mathbf{x}$ indicates integration over $x_1, \ldots, x_m$. Divergence (or *relative entropy*) measures the dissimilarity between two distributions and has many applications in information theory. For our purposes, it captures the personal cost incurred by the agent when he takes the requisite actions to implement some distribution f when he prefers distribution g. Intuitively, the larger the divergence from g to f, the larger the personal cost borne by the agent.

The KL-divergence cost function has several appealing properties which we discuss in Bonham and Riggs-Cragun [2021] and repeat here. First, the cost function is non-negative: $D(f||g) = 0$ for $f = g$ and is positive otherwise. Intuitively, the agent suffers zero 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, Theorem 2.7.2). For a given g, it is increasingly costly for the agent to increase the probability of a particular $\mathbf{x}$, and it is maximally costly to implement a degenerate distribution that *guarantees* the realization of a particular $\mathbf{x}$. This convexity implies that the agent's control over the *distribution* will never amount to full control over the *realization*, because controlling the realization is too costly. In more applied terms, the agent can reduce, but not eliminate, the role of exogenous influences on x. Third, the KL divergence marginal cost approaches infinity as $f(\mathbf{x})$ approaches zero for any $\mathbf{x}$; this will simplify our analysis by guaranteeing interior solutions.

It is useful to think of g as the distribution over $\mathbf{x}$ when the agent exerts minimum effort. For example, g might be the distribution in place when the agent is first hired, perhaps due to actions taken by prior agents. Alternatively, g might be the distribution that would result if the agent spent all his time watching Netflix instead of working. The distribution g is undesirable from the principal's perspective but is maximally desirable from the agent's perspective, *absent incentives*. In the classic parametric model, the principal and agent have a conflict over the agent's effort level, a; the agent prefers minimum effort (e.g. $a = 0$), while the principal wants the agent to work as much as possible. Under the generalized approach, the principal and agent have a conflict of interest over what distribution should be implemented, f. The principal wants an f with the highest possible $\mathbb{E}[B(\mathbf{x})]$, and because she is risk neutral, does not care intrinsically about the distribution's shape. The agent prefers to set $f = g$, and must be provided with incentives to do anything else.

When offered contract s, the agent chooses f to maximize his expected utility from compensation less his personal cost.

$$\begin{aligned} \max_{f} \quad & \int U(s(\mathbf{x}))f(\mathbf{x})d\mathbf{x} - \int f(\mathbf{x}) \ln\left(\tfrac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x} \\ \text{s.t.} \quad & 1 = \int f(\mathbf{x})d\mathbf{x} \\ & f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \end{aligned} \tag{7}$$

The constraints ensure that the chosen f is a p.d.f. Let ν denote the Lagrange multiplier on the constraint $1 = \int f(\mathbf{x})d\mathbf{x}$. We ignore the final set of constraints because, as we will establish shortly, $f(\mathbf{x}) \geq 0$ does not bind for any $\mathbf{x}$. Pointwise optimization of (7) yields the following incentive compatible action.

$$f(\mathbf{x}) \quad = \quad g(\mathbf{x})e^{U(s(\mathbf{x}))-\nu-1}, \tag{8}$$

where $\nu = \ln\left(\int g(\mathbf{x})e^{U(s(\mathbf{x}))-1}d\mathbf{x}\right)$ is obtained by substituting (8) into the constraint $\int f(\mathbf{x})d\mathbf{x} = 1$. Because both g and the exponential function are non-negative everywhere, the unconstrained solution satisfies the constraint $f(\mathbf{x}) \geq 0$ for all $\mathbf{x}$. It follows that when faced with a particular incentive scheme s, the agent chooses f such that (8) is maintained for all $\mathbf{x}$. Taking into account that the agent will respond in this way, the principal solves the following program.

$$\begin{aligned} \max_{s,f} \quad & \int (B(\mathbf{x}) - s(\mathbf{x}))f(\mathbf{x})d\mathbf{x} \\ \text{s.t.} \quad & \nu + \int (U(s(\mathbf{x})) - \nu)f(\mathbf{x})dx - V(f) \geq \bar{U} \\ & U(s(\mathbf{x})) = \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) + 1 + \nu \text{ for all } \mathbf{x} \\ & 1 = \int f(\mathbf{x})d\mathbf{x} \end{aligned} \tag{9}$$

The principal seeks to maximize her expected net payoff subject to three constraints. The first is the IR constraint, where we have added and subtracted ν on the left-hand side for convenience. The second set of constraints are the IC constraints. These are obtained from the agent's first-order condition (equation 8); the first-order approach is valid here because the agent's program (equation 7) maximizes a concave function with linear constraints. There is an IC constraint for every $\mathbf{x}$ because the agent chooses the probability of every possible realization of $\mathbf{x}$, and thus equation (8) must be satisfied for all $\mathbf{x}$ for the proposed (s, f) to be incentive compatible. The final constraint ensures that the agent chooses a distribution that integrates to one; we refer to this as the "p.d.f. constraint."[3] Throughout the paper, we let λ be the Lagrange multiplier on the IR constraint, $\mu(\mathbf{x})$ be the IC multiplier for a given $\mathbf{x}$, and η be the multiplier on the p.d.f. constraint.

Proposition 1 *If the principal's objective is $B(\mathbf{x})$, the agent controls $f(\mathbf{x})$, and $\mathbf{x}$ is available for contracting, the optimal contract is characterized as follows.*

$$\frac{1}{U'(s(\mathbf{x}))} = \lambda - \eta + B(\mathbf{x}) - s(\mathbf{x}). \tag{10}$$

Notice that $\mathbf{x}$ enters (10) only through the contract, $s(\mathbf{x})$, and the objective, $B(\mathbf{x})$. In stark

[3]Equivalently, we could write this constraint as $\nu = \ln\left(\int g(\mathbf{x})e^{U(s(\mathbf{x}))-1}d\mathbf{x}\right)$ to ensure that ν is specified such that $f(\mathbf{x})$ integrates to one in the agent's program. Notice that substituting the IC contract $U(s(\mathbf{x})) = \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) + \nu + 1$ into $\nu = \ln\left(\int g(\mathbf{x})e^{U(s(\mathbf{x}))-1}d\mathbf{x}\right)$ reduces to $1 = \int f(\mathbf{x})d\mathbf{x}$.

contrast to the classic parametric approach, likelihood ratios play no role in the solution, and any variation in the contract comes only through variation in the objective $B(\mathbf{x})$ (see Bonham and Riggs-Cragun [2021] for a detailed analysis and discussion of why this is the case). This has important implications for optimal aggregation.

Rearranging (10) gives

$$s(\mathbf{x}) = \tilde{U}^{-1}(B(\mathbf{x})), \tag{11}$$

where $\tilde{U}(s) \equiv s + \frac{1}{U'(s)} - \lambda + \eta$. As in Banker and Datar [1989], we can separate the solution into two stages, where information in $\mathbf{x}$ is first aggregated into a composite measure and then a compensation contract is written on that aggregate.

$$\begin{array}{rrcl} \textit{Aggregator} & \pi(\mathbf{x}) & \equiv & B(\mathbf{x}) \\ \textit{Contract} & s(\pi(\mathbf{x})) & = & \tilde{U}^{-1}(\pi(\mathbf{x})) \end{array} \tag{12}$$

Thus, the principal's objective, $B(\mathbf{x})$, is an optimal aggregator for contracting. Banker and Datar [1989] show that under the classic approach, optimal aggregators can be linear if likelihood ratios are linear; here, the optimal aggregator is linear if the principal's objective is linear. More notably, the optimal weights in the linear aggregator are determined entirely by $B(\mathbf{x})$, as shown in the following corollary.

Corollary 1 *Assume the principal values* $\mathbf{x}$ *linearly so that her objective is* $B(\mathbf{x}) = \mathbf{b}^T\mathbf{x} = \sum_{i=1}^{m} b_i x_i$*, where* $\mathbf{b} \equiv (b_1 \ldots b_m)^T$*. Then an optimal linear aggregator and contract solving the principal's program are as follows.*

$$\begin{array}{rrcl} \text{Aggregator} & \pi(\mathbf{x}) & \equiv & \sum_{i=1}^{m} \pi_i x_i = \sum_{i=1}^{m} b_i x_i \\ \text{Contract} & s(\pi(\mathbf{x})) & = & \tilde{U}^{-1}(\pi(\mathbf{x})), \end{array} \tag{13}$$

where $\tilde{U}(s) \equiv s + \frac{1}{U'(s)} - \lambda + \eta$ *characterizes the contract's functional form. In particular, if* $B(\mathbf{x})$ *is a linear aggregate, then the contract is conditioned on that same linear aggregate, with the optimal weight on* x_i *given by* b_i*.*

The corollary shows that under the generalized distribution approach, the optimal linear aggregator simply sets $\pi_i = b_i$ for each x_i. That is, performance measures are weighted according to their weights in the principal's objective. Sensitivity and precision (however defined in this setting) play no role in the aggregation process; the *only* thing that matters for optimal weighting is the vector $\mathbf{b}$. The optimal linear aggregator $\pi(\mathbf{x})$ weights two measures x_i and x_j by their proportion in the principal's objective.

$$\frac{\pi_i}{\pi_j} = \frac{b_i}{b_j} \tag{14}$$

Corollary 1 helps to rationalize the use of equally-weighted aggregates in executive compensation contracts. As an example, assume that $\mathbf{x} = (x_r, x_e)^T$, where x_r is total revenues and x_e is total expenses; and assume that the principal's objective is $B(x_r, x_e) = x_r - x_e$. Assume that revenues are noisier than expenses because, for example, demand fluctuates but salaries are fixed. The classic sensitivity-precision result from Banker and Datar [1989] would predict a higher weight on expenses than revenues all else equal, and therefore it would be inefficient to contract on earnings, $z \equiv x_r - x_e$, an *equally*-weighted difference between revenues and expenses. Under our approach, by contrast, the optimal aggregator follows from (13) as

$$\pi(x_r, x_e) = B(x_r, x_e) = x_r - x_e = z. \tag{15}$$

That is, contracting on net income is a perfectly efficient, despite the noisiness of revenues relative to expenses.

An obvious shortcoming of this example is that accounting measures such as GAAP revenues and expenses are not what firm owners value *per se*, but instead are *estimates* of underlying economic phenomena that firm owners do care about. If firm owners care about "true revenues" and "true expenses" (which are unobservable), can GAAP net income still be an efficient aggregator? We address this question in the next section.

4 Contracting on Aggregated Estimates

Until this point in our analysis, we have assumed that the components of the principal's objective, $x_1, \ldots, x_m$, are all observable and contractible. But this assumption is inconsistent with the way accountants typically think. If the things firm owners value were contractible, there would be no need for measurement. It is common practice in accounting theory to assume that firm owners care about *firm fundamentals*, and that these underlying economic phenomena are measured or estimated by the accounting process. To better capture the accounting process, in this section we extend our model to a setting where the things firm owners care about are unobservable.

Assume as before that the principal cares about a set of m random variables given by $\mathbf{x} \equiv (x_1, \ldots, x_m)^T$, and assume that she values these measures linearly so that her objective is $B(\mathbf{x}) = \sum_{i=1}^{m} b_i x_i$. Assume that $\mathbf{x}$ is unobservable, but that the principal has at her disposal a set of n contractible random variables given by $\mathbf{y} \equiv (y_1, \ldots, y_n)^T$. Let the agent choose $f(\mathbf{x}, \mathbf{y}) \in \Delta(\mathbf{x}, \mathbf{y})$, where $\Delta(\mathbf{x}, \mathbf{y})$ is the space of all joint probability distributions over $\mathbf{x}$ and $\mathbf{y}$. We might think of $\mathbf{x}$ as including constructs such as the change in true firm value during the contracting period or the value of assets in place, while $\mathbf{y}$ might represent observable detailed data such as individual sales transactions, inventory delivery times, amounts spent on R&D projects, and the historical purchase prices of property, plant and equipment. We define the cost of implementing $f(\mathbf{x}, \mathbf{y})$ by its divergence from an exogenous reference distribution,

$g(\mathbf{x},\mathbf{y}) \in \Delta(\mathbf{x},\mathbf{y})$:

$$V(f) \;=\; D(f(\mathbf{x},\mathbf{y})||g(\mathbf{x},\mathbf{y})) \;\equiv\; \int f(\mathbf{x},\mathbf{y}) \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) d(\mathbf{x},\mathbf{y}), \tag{16}$$

where $\int d(\mathbf{x},\mathbf{y})$ indicates integration over $x_1,\ldots,x_m,y_1,\ldots,y_n$. As before, $g(\mathbf{x},\mathbf{y})$ minimizes $V(f)$ and is therefore the distribution preferred by the agent absent incentives. Given a contract $s(\mathbf{y})$, the agent chooses $f(\mathbf{x},\mathbf{y})$ to maximize his expected utility from compensation minus his personal cost.

$$\begin{aligned} \max_{f} \quad & \int U(s(\mathbf{y}))f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) - \int \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & 1 = \int f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{17}$$

Letting ν be the Lagrange multiplier on the constraint, pointwise optimization yields the following characterization of the incentive compatible action.

$$f(\mathbf{x},\mathbf{y}) \;=\; g(\mathbf{x},\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu} \tag{18}$$

The non-contractability of $\mathbf{x}$ causes the agent to have more control in choosing a distribution than the principal has in designing the contract. The agent chooses the probability of every $(\mathbf{x},\mathbf{y})$ while the principal chooses a payment $s(\mathbf{y})$ for every outcome $\mathbf{y}$. The principal is faced with the following optimization problem.

$$\begin{aligned} \max_{s,f} \quad & \int (B(\mathbf{x}) - s(\mathbf{y}))f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \nu + \int \left(U(s(\mathbf{y})) - \nu\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) - \int \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \geq \bar{U} \\ & U(s(\mathbf{y})) = \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) + 1 + \nu \text{ for all } (\mathbf{x},\mathbf{y}) \\ & 1 = \int f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{19}$$

Due to the principal's control disadvantage, there are some distributions that she cannot implement with any contract.[4] Pointwise optimization of program (19) over s is done pointwise at every $\mathbf{y}$ (rather than at every $(\mathbf{x}, \mathbf{y})$), which leaves integrals over $\mathbf{x}$ in the solution. In particular, the optimal contract depends on conditional expectations over $\mathbf{x}$ given $\mathbf{y}$. This is shown in the following proposition, which presents the solution to program (19).

Proposition 2 *If the agent chooses $f(\mathbf{x}, \mathbf{y})$ nonparametrically, where the principal's objective is $B(\mathbf{x}) = \sum_{i=1}^{m} b_i x_i$ but she can only contract on $\mathbf{y}$, the optimal contract is characterized as follows.*

$$\begin{aligned} \frac{1}{U'(s(\mathbf{y}))} &= \lambda - \eta + \int B(\mathbf{x}) f(\mathbf{x}|\mathbf{y}) d\mathbf{x} - s(\mathbf{y}) \\ \iff s(\mathbf{y}) &= \tilde{U}^{-1}\left(\mathbb{E}_f[B(\mathbf{x})|\mathbf{y}]\right), \end{aligned} \tag{20}$$

where $\mathbb{E}_f[B(\mathbf{x})|\mathbf{y}] \equiv \int B(\mathbf{x}) f(\mathbf{x}|\mathbf{y}) d\mathbf{x}$ is the expected value of $B(\mathbf{x})$ given $\mathbf{y}$ under the equilibrium distribution f, and $\tilde{U}(s) \equiv s + \frac{1}{U'(s)} - \lambda + \eta$.

The proposition shows that the optimal contract is a transformation of $\mathbb{E}_f[B(\mathbf{x})|\mathbf{y}]$, the expected value of the principal's objective given all available information in $\mathbf{y}$. Because we have assumed $B(\mathbf{x})$ to be linear, $\mathbb{E}_f[B(\mathbf{x})|\mathbf{y}]$ is a linear sum of conditional expectations:

$$\mathbb{E}_f[B(\mathbf{x})|\mathbf{y}] = \mathbb{E}_f\left[\sum_{i=1}^{m} b_i x_i \middle| \mathbf{y}\right] = \sum_{i=1}^{m} b_i \, \mathbb{E}_f[x_i|\mathbf{y}], \tag{21}$$

where $\mathbb{E}_f[x_i|\mathbf{y}] \equiv \int x_i f(x_j|\mathbf{y}) dx_i$ is the expected value of x_i given all available information in $\mathbf{y}$. The following corollary shows that we can separate the principal's solution into three components: estimation, aggregation, and compensation. We let $\hat{x}_i$ denote the principal's estimate of x_i and $\hat{\mathbf{x}}$ denote her estimate of the vector $\mathbf{x}$.

Corollary 2 *If the agent chooses the distribution $f(\mathbf{x}, \mathbf{y})$ where the principal's objective is $B(\mathbf{x}) = \sum_{i=1}^{m} b_i x_i$ and $\mathbf{y}$ is the set of contractible performance measures, then an optimal*

[4]Specifically, the principal can only implement contracts for which $f(\mathbf{x}|\mathbf{y}) = g(\mathbf{x}|\mathbf{y})$ for all $(\mathbf{x}, \mathbf{y})$ (see Lemma 3 of Bonham and Riggs-Cragun [2021]).

estimate, aggregator, and compensation contract are as follows.

$$\begin{array}{lrcl} \text{Estimate} & \hat{\mathbf{x}} & = & \mathbb{E}_f[\mathbf{x}|\mathbf{y}] \\ \text{Aggregator} & \pi(\hat{\mathbf{x}}) & = & B(\hat{\mathbf{x}}) \equiv \mathbf{b}^T\hat{\mathbf{x}} \\ \text{Contract} & s(\pi(\hat{\mathbf{x}})) & = & \tilde{U}^{-1}\left(\pi(\hat{\mathbf{x}})\right), \end{array} \tag{22}$$

where $\mathbb{E}_f[\mathbf{x}|\mathbf{y}]$ *is the expected value of* $\mathbf{x}$ *given* $\mathbf{y}$ *under the equilibrium distribution* f*;* $\pi(\hat{\mathbf{x}})$ *is the optimal aggregated estimate; and* $\tilde{U}(s) \equiv \frac{1}{U'(s)} + s - \lambda + \eta$ *characterizes the agent's compensation function.*

The solution is executed in three stages. First, the principal estimates each x_j as $\hat{x}_j = \mathbb{E}_f[x_j|\mathbf{y}] \equiv \int x_j f(x_j|\mathbf{y})dx_j$. Notice that $\hat{x}_j$ is an *unbiased estimate*, or in the words of accounting standard setters, it is a "faithful representation of the real-world economic phenomena that it purports to represent" (FASB [2006]). Second, the principal aggregates the estimates according to the weights in her objective to produce the aggregated estimate $\pi(\hat{\mathbf{x}}) = \mathbf{b}^T\hat{\mathbf{x}} \equiv \sum_{i=1}^{m} b_i\hat{x}_i$. Finally, the principal conditions the agent's compensation on this aggregated estimate.

To illustrate, let us return to the example posited in section 3, where the principal cares about true revenues, x_r, and true expenses, x_e, such that her objective is $B(x_r, x_e) = x_r - x_e \equiv z$, where z denotes true earnings. Now suppose that these true values are unobservable, but that the principal does observe $\mathbf{y}$, which contains information such as sales transactions, customer characteristics, inventory data, and operational expenditures. One option is that the principal could condition the contract directly on this underlying data (indeed, compensation committees have enormous databases of detailed information at their fingertips). Corollary 2, provides a different option. The principal first uses information in $\mathbf{y}$ to construct unbiased estimates of revenues and expenses, $\hat{x}_r$ and $\hat{x}_e$:

$$\begin{aligned} \textit{Estimate of } x_r \qquad \hat{x}_r &= \mathbb{E}_f[x_r|\mathbf{y}] \\ \textit{Estimate of } x_e \qquad \hat{x}_e &= \mathbb{E}_f[x_e|\mathbf{y}]. \end{aligned} \tag{23}$$

Next, the principal aggregates these estimates according to the weights in her objective. Because we have assumed that $B(x_r, x_e)$ is an equally weighted difference between revenues and expenses, the principal can efficiently condition the agent's compensation on net income, $\hat{z} \equiv \hat{x}_r - \hat{x}_e$.

$$\begin{aligned} \textit{Aggregator} \qquad \pi(\hat{x}_r, \hat{x}_e) &= B(\hat{x}_r, \hat{x}_e) = \hat{x}_r - \hat{x}_e \equiv \hat{z} \\ \textit{Contract} \qquad s(\pi(\hat{x}_r, \hat{x}_e)) &= s(\hat{z}) = \tilde{U}^{-1}(\hat{z}). \end{aligned} \tag{24}$$

Then to the extent that GAAP revenues and expenses are representationally faithful, contracting solely on net income can be just as efficient as writing an extremely complex contract conditioned on all of the underlying information used to construct net income.

Our results can also provide a rationale for contracting on both revenue and income metrics, a commonly observed practice in executive compensation (De Angelis and Grinstein [2015], Bloomfield, Gipper, Kepler, and Tsui [2021]). Extending the example above, suppose that the principal cares more about revenues than she does about expenses, such that her objective is given by $B(x_r, x_e) = b_r x_r - x_e$, with $b_r > 1$. In this case, the optimal contract can be conditioned on a weighted aggregate of estimated revenues and expenses,

$$\pi(\hat{x}_r, \hat{x}_e) = b_r \hat{x}_r - \hat{x}_e, \tag{25}$$

or equivalently, on revenues and net income:

$$\pi(\hat{x}_r, \hat{z}) = (b_r - 1)\hat{x}_r + \hat{z}, \tag{26}$$

where $b_r - 1 > 0$. Thus, contracts that put positive incentives in revenues and earnings are in effect weighting revenues more heavily than expenses. This equivalence is also pointed out by Bloomfield et al. [2021], who interpret the practice as a form of "cost shielding." They document that CEO bonus plans include an average of 2.27 income statement measures, where 0.5 are revenue metrics, 1.01 are net income, and the remainder are earnings measures other than net income (e.g. EBITDA, EBIT, and EBT). Based on our results, these statistics suggest that shareholders tend to care more about revenues than about expenses. By contrast, applying the findings from Banker and Datar [1989] to these statistics would suggest that revenues tend to be less noisy than expenses or more sensitive to CEO actions.

5 Optimal Measurement

Corollary 2 shows that the optimal contract is conditioned on the unbiased estimates $\hat{\mathbf{x}} = \mathbb{E}_f[\mathbf{x}|\mathbf{y}]$. We refer to accounting *measurement* as the process by which the principal forms the estimates $\hat{\mathbf{x}}$. Under the general model in the prior section, the measurement process is ambiguous; without knowing the form of f, we can't say anything about how the principal forms the expectation $\mathbb{E}_f[\mathbf{x}|\mathbf{y}]$. In this section, we employ a simple specification of our model in which the optimal $f(\mathbf{x}, \mathbf{y})$ arising in equilibrium is a multivariate normal distribution. This gives us the tractability to open the measurement black box.

As in section 4, assume that the principal's objective is $B(\mathbf{x}) = \mathbf{b}^T\mathbf{x}$, that $\mathbf{y}$ is contractible but $\mathbf{x}$ is not, and that the agent controls $f(\mathbf{x}, \mathbf{y})$ at a personal cost given by (16). Assume that the agent's cost-minimizing distribution, $g(\mathbf{x}, \mathbf{y})$, is a multivariate normal distribution with mean $\mathbf{0}$ and variance-covariance matrix $\Sigma_g \equiv \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$. Finally, assume that the agent's reservation utility is $\bar{U} = 0$, and assume that he is risk neutral so that $U(s) = s$. We are giving up risk aversion for tractability here, and we feel this is a worthy sacrifice given that this section is focused on measurement rather than the shape of the compensation function.

The following proposition provides the optimal contract and shows that the agent's equilibrium action mean-shifts the distribution g.

Proposition 3 *Assume that the principal's objective is $B(\mathbf{x}) = \mathbf{b}^T\mathbf{x}$, but the only measure available for contracting is $\mathbf{y}$. Let the risk-neutral agent choose $f(\mathbf{x}, \mathbf{y})$, and assume that $g(\mathbf{x}, \mathbf{y})$ is a centered multivariate normal distribution with covariance matrix $\Sigma_g \equiv \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$. Then an optimal estimate, aggregator, contract, and action are characterized as follows.*

$$\begin{array}{lrcl} \text{Estimate} & \hat{\mathbf{x}} & = & \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y} \\ \text{Aggregator} & \pi(\hat{\mathbf{x}}) & = & B(\hat{\mathbf{x}}) = \mathbf{b}^T\hat{\mathbf{x}} \\ \text{Contract} & s(\pi(\hat{\mathbf{x}})) & = & \pi(\hat{\mathbf{x}}) - \eta \\ \text{Action} & f(\mathbf{x}, \mathbf{y}) & = & g(\mathbf{x}, \mathbf{y})e^{\mathbf{b}^T\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y}-\eta}, \end{array} \tag{27}$$

where $\eta = \frac{1}{2}\mathbf{b}^T\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{b}$. Under the equilibrium distribution, $(\mathbf{x}, \mathbf{y}) \sim \mathcal{N}\left(\left[\begin{smallmatrix}\mu_{\mathbf{x}} \\ \mu_{\mathbf{y}}\end{smallmatrix}\right], \Sigma_g\right)$, where $\mu_{\mathbf{x}} = \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{b}$ and $\mu_{\mathbf{y}} = \Sigma_{\mathbf{yx}}\mathbf{b}$.

Because we are interested in measurement, the solution (27) is presented using a three-stage decomposition, as in Corollary 2, where the principal first constructs the estimates $\hat{x}_1 \ldots \hat{x}_m$, aggregates them according to the weights in her objective, and then conditions the agent's compensation on the aggregated estimate. The three-step solution is equivalent to contracting directly on $\mathbf{y}$, where $s(\mathbf{y}) = \mathbf{b}^T\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y}$.

Our specification in this section changes the general solution (22) in three ways. First, measurement is no longer ambiguous; the principal's estimate is given by the linear regression $\hat{\mathbf{x}} = E_f[\mathbf{x}|\mathbf{y}] = \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y}$.[5] Second, the optimal contract is linear in the estimates $\hat{\mathbf{x}}$. This linearity comes from assuming risk neutrality. While risk neutrality results in linear contracts

[5]The measurement process will always have $\hat{\mathbf{x}} = \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y}$ when g is a multivariate normal, even if the agent is not risk neutral. This is because $f(\mathbf{x}|\mathbf{y}) = g(\mathbf{x}|\mathbf{y})$ always holds when $\mathbf{x}$ is not contractible (see Bonham and Riggs-Cragun [2021], Lemma 3), so $f(\mathbf{x}|\mathbf{y})$ is a normal distribution even if $f(\mathbf{x}, \mathbf{y})$ is not. Thus, $\mathbb{E}_f[\mathbf{x}|\mathbf{y}]$ is a linear regression of $\mathbf{x}$ on $\mathbf{y}$ for any risk preference, and risk neutrality is needed only for normality in $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}|\mathbf{y})f(\mathbf{y})$.

being weakly optimal under the classic approach, here it results in linear contracts that are *uniquely* optimal. Third, we now have a closed-form solution for the equilibrium f. The proposition shows that if g is a multivariate normal with mean $\mathbf{0}$ and variance-covariance Σ_g, then the equilibrium f is also a multivariate normal with different means but the same variance-covariance matrix as g.

It is worth emphasizing that we have not restricted f to be normal; as in our general model, the agent has the ability to implement *any* distribution imaginable. The reason a normal distribution arises in equilibrium has to do with the KL-divergence cost function, in combination with the assumption of risk neutrality. The risk-neutral principal does not care about the shape of the distribution and wants to increase its mean as cheaply as possible. It can be shown that for a given mean shift from g to f where g is a normal distribution with variance-covariance matrix Σ_g, the f that minimizes KL divergence is a normal distribution with variance-covariance matrix Σ_g. Combining this with a risk-neutral agent means that making f a normal distribution is the cheapest way to implement a given mean shift.

The normality of f conveniently allows for comparative statics on how Σ_g affects production. As a simple example, consider the case with $m = n = 1$; that is, the principal cares about a single variable and only one transaction characteristic is contractible. Specifically, assume $B(\mathbf{x}) = x$ and $\mathbf{y} = y$, where the agent chooses $f(x, y)$. Assume $g(x, y)$ is a centered bivariate normal distribution with correlation coefficient $\rho > 0$, so that $\Sigma_g = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$. Then the optimal estimate, contract and distribution are given as follows.

$$\begin{array}{rrcl} \textit{Estimate} & \hat{x} & = & \rho y \\ \textit{Contract} & s(\hat{x}) & = & \hat{x} - \frac{1}{2}\rho^2 \\ \textit{Action} & f(x,y) & \sim & \mathcal{N}\left(\left[\begin{smallmatrix} \rho^2 \\ \rho \end{smallmatrix}\right], \Sigma_g\right) \end{array} \tag{28}$$

In equilibrium, $\mu_x = \rho^2$ and $\mu_y = \rho$. Notice that $\rho^2 < \rho$ for $0 < \rho < 1$, so in equilibrium the agent mean-shifts y more than x. The parameter ρ indexes the complementarity between x and

y in the agent's preferred distribution g. When ρ is small, it is easy for the agent to improve y without improving x. When ρ is high, the agent finds it costly to move y without also moving x. When $\rho = 1$, x and y are *perfect* complements in the agent's preferred distribution, and he prefers to move x and y in exactly the same way; in this case, the principal does just as well contracting on y as she would do if she could contract on x directly.

We will explore the $m = n = 1$ case further in section 5.1, which is the first of two accounting measurement applications that we provide. Application 5.1 studies how optimal measurement is affected by the agent's ability to manage earnings, and application 5.2 studies optimal measurement of uncertain investments. These applications are highly stylized and should be taken as suggestive examples; we leave it to future work to develop more rigorous analyses of these issues. The applications do suggest that many common accounting practices are efficient for contracting when managers have extensive control over both fundamental performance and accounting reports.

5.1 Window dressing and conservatism

In many settings, managers can engage in *window dressing* actions – non-value-added activities that improve a performance measure but do not contribute to the principal's objective. Our model is well-suited for studying this issue. Let the principal's objective be true performance, denoted x, and assume that x is not contractible. Let y be contractible evidence about the realization of x, such as a transaction characteristic as in Gao [2013]. The agent controls $f(x, y)$, and thus can potentially engage in window dressing by manipulating the transaction characteristic to improve the distribution of y. The principal must design a measurement rule $\hat{x}(y)$ and contract $s(\hat{x})$ to optimize her net payoff in the face of the agent's ability to game y.

Assume that the agent's preferred distribution, $g(x, y)$, is a centered bivariate normal with unit variance and correlation coefficient ρ. If ρ is close to one, the agent finds it very costly to increase y without also increasing x; that is, window dressing behavior is difficult and the transaction characteristic provides very reliable evidence about x. By contrast, if ρ is close to zero, the agent finds it very easy to increase y without changing x; that is, window dressing behavior is very easy and the transaction characteristic provides little reliable evidence.

Define window dressing as the extent to which the expected transaction characteristic exceeds expected true performance, $\mathbb{E}_f[y] - \mathbb{E}_f[x]$. The solution is given by equation (28) and shows that the amount of window dressing that takes place in equilibrium is $\mu_x - \mu_y = \rho - \rho^2$. This function is concave with a global maximum at $\rho = \frac{1}{2}$ and is equal to zero at $\rho = 0$ or $\rho = 1$. Thus, for any interior level of susceptibility to window dressing (i.e. $\rho \in (0, 1)$), there will be some amount of window dressing activity in equilibrium.

The principal is not fooled by the agent's window dressing activity, and she takes it into account when forming her estimate. By (28), the optimal measurement rule sets $\hat{x} = \rho y$, which implies that $\mathbb{E}_f[\hat{x}] = \rho\,\mathbb{E}_f[y] = \rho^2 = \mathbb{E}_f[x]$. That is, the optimal measurement rule discounts y by factor $\rho \in [0, 1]$ in order to reduce $\mathbb{E}_f[\hat{x}]$ from $\mathbb{E}_f[y]$ to $\mathbb{E}_f[x]$. As a result, the optimal estimate $\hat{x}$ *understates* the manipulable evidence y but is an unbiased measure of the true value of x. Analogous to Gao [2013] and as suggested by Watts [2003], this *conservative measurement rule* offsets managerial biases to create an *unbiased performance measure*.

5.2 Accounting for investments

Let the principal's objective be $B(\mathbf{x}) = b_2 x_2 - x_1$, where x_1 represents current outlays in an investment and x_2 represents future benefits from the investment. Let y_1 and y_2 represent evidence about x_1 and x_2, respectively. The agent chooses the joint distribution $f(x_1, x_2, y_1, y_2)$. Let the agent's cost-minimizing distribution, $g(x_1, x_2, y_1, y_2)$, be a centered multivariate normal with variance-covariance matrix

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{x_1x_2} & \rho_{x_1y_1} & \rho_{x_1y_2} \\ \rho_{x_1x_2} & 1 & \rho_{x_2y_1} & \rho_{x_2y_2} \\ \rho_{x_1y_1} & \rho_{x_2y_1} & 1 & \rho_{y_1y_2} \\ \rho_{x_1y_2} & \rho_{x_2y_2} & \rho_{y_1y_2} & 1 \end{bmatrix}. \tag{29}$$

We assume that current investment outlays are perfectly measurable by setting $\rho_{x_1y_1} = 1$. This assumption implies that $\rho_{x_1z} = \rho_{y_1z}$ for any measure z, and consequently, we have $\rho_{x_2y_1} = \rho_{x_1x_2} \equiv \rho_x$ and $\rho_{x_1y_2} = \rho_{y_1y_2} \equiv \rho_y$. Let $\rho_2 \equiv \rho_{x_2y_2}$. Then the variance-covariance matrix above can be rewritten as

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_x & 1 & \rho_y \\ \rho_x & 1 & \rho_x & \rho_2 \\ 1 & \rho_x & 1 & \rho_y \\ \rho_y & \rho_2 & \rho_y & 1 \end{bmatrix}. \tag{30}$$

From Proposition 3, the solution is as follows.

$$\begin{aligned}
\textit{Estimate of } x_1\text{:} \quad & \hat{x}_1 = y_1 \\
\textit{Estimate of } x_2\text{:} \quad & \hat{x}_2 = \left(\tfrac{\rho_x - \rho_y\rho_2}{1-\rho_y^2}\right) y_1 + \left(\tfrac{\rho_2 - \rho_x\rho_y}{1-\rho_y^2}\right) y_2 \\
\textit{Aggregator}\text{:} \quad & \pi(\hat{\mathbf{x}}) = B(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 \\
\textit{Contract}\text{:} \quad & s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\
\textit{Action}\text{:} \quad & f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = \rho_x b_2 - 1 \\ \mu_{x_2} = \frac{b_2(\rho_x^2 + \rho_2^2 - 2\rho_x\rho_y\rho_2) - \rho_x(1-\rho_y^2)}{1-\rho_y^2} \\ \mu_{y_1} = \rho_x b_2 - 1 \\ \mu_{y_2} = \rho_2 b_2 - \rho_y \end{bmatrix}, \Sigma_g \right)
\end{aligned} \tag{31}$$

To gain intuition for this solution, we will do comparative statics under two cases, one in which reliable evidence about future benefit x_2 is very difficult to produce ($\rho_2 = 0$) and one where it is very easy to produce ($\rho_2 = 1$).

Case 1: Future returns difficult to measure

We first consider the case in which it is very difficult to produce reliable evidence about future returns on investment. We capture this setting by assuming that, absent incentives, y_2 is pure white noise: $\rho_{x_1 y_2} = \rho_{x_2 y_2} = \rho_{y_1 y_2} = 0$. With this assumption, the variance-covariance matrix under the agent's cost-minimizing distribution g is given as follows, where $\rho_x \equiv \rho_{x_1 x_2} = \rho_{x_2 y_1}$.

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_x & 1 & 0 \\ \rho_x & 1 & \rho_x & 0 \\ 1 & \rho_x & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{32}$$

Under these assumptions, the solution (31) reduces to the following.

$$\begin{aligned} \textit{Estimate of } x_1\text{:} \quad & \hat{x}_1 = y_1 \\ \textit{Estimate of } x_2\text{:} \quad & \hat{x}_2 = \rho_x y_1 = \rho_x \hat{x}_1 \\ \textit{Aggregator:} \quad & \pi(\hat{\mathbf{x}}) = b_2 \hat{x}_2 - \hat{x}_1 = (b_2\rho_x - 1)\hat{x}_1 \\ \textit{Contract:} \quad & s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\ \textit{Action:} \quad & f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = b_2\rho_x - 1 \\ \mu_{x_2} = \rho_x(b_2\rho_x - 1) \\ \mu_{y_1} = b_2\rho_x - 1 \\ \mu_{y_2} = 0 \end{bmatrix}, \Sigma_g \right) \end{aligned} \tag{33}$$

Both $\hat{x}_1$ and $\hat{x}_2$ are unbiased measures of x_1 and x_2 given the equilibrium action: $\mathbb{E}_f[\hat{x_1}] = \mathbb{E}_f[y_1] = b_2\rho_x - 1 = \mu_{x_1}$ and $\mathbb{E}_f[\hat{x_2}] = \rho_x \mathbb{E}_f[y_1] = \mu_{x_2}$. Interpret the aggregation function $\pi(\hat{\mathbf{x}}) = (b_2\rho_x - 1)\hat{x}_1$ as reported net income, and note that this is increasing in ρ_x for all positive investments.

If $\rho_x = 0$ so that investments are maximally uninformative about future returns, then net income is given by $\pi(\hat{\mathbf{x}}) = -\hat{x}_1$. That is, estimated investments are all immediately expensed and no future benefits are estimated. This is loosely analogous to the treatment of R&D, where investments have a very low correlation with future returns and are immediately expensed. Given that net income is constructed in this way when $\rho_x = 0$, the agent chooses a distribution in which $\mu_{x_2} = 0$ and $\mu_{x_1} = -1$; that is, the agent ignores future returns and liquidates some existing projects (i.e., cuts R&D) to make a short-term profit.

As ρ_x increases from zero, a smaller proportion of $\hat{x}_1$ is optimally deducted from net income. When $\rho_x = \frac{1}{b_2}$, net income is equal to $\pi(\hat{\mathbf{x}}) = (b_2\rho_x - 1)\hat{x}_1 = 0$. That is, estimated investments $\hat{x}_1$ are never reflected on the income statement; i.e., they are capitalized. We can interpret $\rho_x \in \left(0, \frac{1}{b_2}\right)$ as investments that are partly capitalized and partly expensed. This is loosely analogous to investments in fixed assets, which indeed are capitalized rather than expensed, with some amount deducted from net income as depreciation. For fixed asset investments with a relatively high likelihood of return, i.e. $\rho_x = \frac{1}{b_2}$, the agent chooses a distribution in which $\mu_{x_1} = \mu_{x_2} = 0$; that is, the agent maintains the status quo investment strategy and refrains from liquidating PP&E.

Finally, as ρ_x increases beyond $\frac{1}{b_2}$, net income optimally includes some unrealized gains or revenues; in the limit where $\rho_x = 1$, net income is given by $\pi(\hat{\mathbf{x}}) = (b_2 - 1)\hat{x}_1$. For instance, the sacrifice of inventory in a credit sale tends to be highly correlated with the future receipt of cash from customers. In this analogy, x_1 is total cost of goods sold, b_2 is the sales price, and $b_2\rho_x$ is the net realizable value of a representative sale. If bad debts are immaterial when the agent does not exert effort, then in equilibrium the agent chooses a distribution in which $\mu_{x_1} = \mu_{x_2} = b_2 - 1 > 0$; that is, the agent exerts effort to sell inventory to customers with good credit.

Case 2: Future returns easy to measure

Now we relax the assumption that y_2 is pure noise and consider how the solution (31) changes under the opposite assumption; that is, as evidence about future payoffs becomes perfectly reliable. Notice that as ρ_2 approaches 1, ρ_x approaches ρ_y.[6] Then setting $\rho_2 = 1$ and letting $\rho \equiv \rho_x = \rho_y$, the variance-covariance matrix is as follows.

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho & 1 & \rho \\ \rho & 1 & \rho & 1 \\ 1 & \rho & 1 & \rho \\ \rho & 1 & \rho & 1 \end{bmatrix}. \tag{34}$$

This reduces solution (31) to the following.

$$\begin{aligned} \textit{Estimate of } x_1&: \quad \hat{x}_1 = y_1 \\ \textit{Estimate of } x_2&: \quad \hat{x}_2 = y_2 \\ \textit{Aggregator}&: \quad \pi(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 \\ \textit{Contract}&: \quad s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\ \textit{Action}&: \quad f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = \rho b_2 - 1 \\ \mu_{x_2} = b_2 - \rho \\ \mu_{y_1} = \rho b_2 - 1 \\ \mu_{y_2} = b_2 - \rho \end{bmatrix}, \Sigma_g \right) \end{aligned} \tag{35}$$

Because the agent in this case has no incentive to window dress y_1 or y_2, the principal takes both signals as given when estimating x_1 and x_2; unlike in section 5.1, conservatism is not needed to construct unbiased estimates. Recall that the principal's objective is $B(\mathbf{x}) = b_2x_2 - x_1$, and suppose that b_2 represents the principal's discount factor on the future cash flow x_2. The solution above gives that reported income is $\pi(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 = b_2y_2 - y_1$. Then the change in

[6] For the variance-covariance matrix of the three random variables x_2, y_1 and y_2 to be positive semi-definite, $\rho_2\rho_y - \sqrt{(1-\rho_2^2)(1-\rho_y^2)} \leq \rho_x \leq \rho_2\rho_y + \sqrt{(1-\rho_2^2)(1-\rho_y^2)}$. Thus, as ρ_2 converges to 1, the lower bound and upper bound on ρ_x both converge to ρ_y.

firm wealth from the investment is reported at *fair value*, expected future cash flows less the amount expended in the current period.

Taken together, the results in this section suggest that optimal measurement rules are driven by the reliability of available evidence. When evidence about future returns is completely unreliable (Case 1), optimal measurement rules are driven by ρ_x, the correlation between current investments and future returns; that is, whether income from a particular investment should be accrued or deferred depends on the likelihood that the investment will pay off in the future. When evidence about future returns becomes available, optimal measurement rules take that evidence into account (see $\hat{x}_2$ in equation 31), and as that evidence becomes *perfectly* reliable, the optimal measurement rule is to report fair value, regardless of the correlation between investment and future returns.

6 Conclusion

The classic sensitivity-precision result from Banker and Datar [1989] shows that performance measures are optimally aggregated according to their relative sensitivities and precision, implying that equally-weighted accounting aggregates are suboptimal for contracting. We revisit optimal aggregation in a model where the agent has intricate control over the distribution of performance measures, and we show that performance measures should be weighted by their weights in the principal's objective. This result indicates a sound reason to revisit empirical studies of executive compensation that rely on the classic sensitivity-precision result. To the extent that executives are better described as influencing a distribution's form rather than its parameters, the generalized approach may have more predictive power than the classic parametric approach. Our results also suggest that empiricists can use executive compensation contracts to make inferences about shareholder or board priorities; specifically, the relative weight on a measure in a CEO's compensation contract should reflect how much shareholders care about that measure relative to other measures in the contract.

Our results depart from Banker and Datar [1989] because we use the *generalized distribution formulation* of the agency problem rather than the conventional *parameterized distribution formulation*. Future theoretical work could provide a reconciliation between the two, perhaps by showing the latter as a special case of the former, or the former as a limiting case of the latter. A reconciliation of this sort would facilitate a sharper comparison of findings across the two approaches, and may generate insights that deepen our understanding of agency problems.

Appendix: Proofs

Proof of Proposition 1. First, write program (9) in Lagrangian form as follows.

$$\begin{aligned}\mathcal{L} &= \int (B(\mathbf{x}) - s(\mathbf{x}))f(\mathbf{x})d\mathbf{x} \\ &+\lambda\left[\nu + \int (U(s(\mathbf{x})) - \nu)f(\mathbf{x})dx - V(f) - \bar{U}\right] \\ &+ \int \mu(\tilde{\mathbf{x}})\left[U(s(\tilde{\mathbf{x}})) - \ln\left(\frac{f(\tilde{\mathbf{x}})}{g(\tilde{\mathbf{x}})}\right) - 1 - \nu\right]d\tilde{\mathbf{x}} \\ &+\eta\left[1 - \int f(\mathbf{x})d\mathbf{x}\right]\end{aligned} \tag{36}$$

Taking the first-order condition with respect to $s(\mathbf{x})$ and rearranging gives

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

Now taking the first-order condition with respect to $f(\mathbf{x})$ gives

$$0 = B(\mathbf{x}) - s(\mathbf{x}) + \lambda\left[U(s(\mathbf{x})) - \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) - 1 - \nu\right] - \mu(\mathbf{x})\frac{1}{f(\mathbf{x})} - \eta \tag{38}$$

Notice that the IC constraint implies that the term in brackets is equal to zero. Then rearranging (38) gives $\mu(\mathbf{x}) = f(\mathbf{x})\left(B(\mathbf{x}) - s(\mathbf{x}) - \eta\right)$. Substituting this into (37) produces the solution presented in the proposition.

□

Proof of Proposition 2. Rearranging (18) gives $U(s(\mathbf{y})) = \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) + 1 + \nu$, allowing us to write the principal's program as follows.

$$
\begin{aligned}
\max_{s,f} \quad & \int (B(\mathbf{x}) - s(\mathbf{y})) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \\
\text{s.t.} \quad & \nu + \int \left(U(s(\mathbf{y})) - \nu\right) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) - \int \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \geq \bar{U} \\
& U(s(\mathbf{y})) = \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) + 1 + \nu \text{ for all } (\mathbf{x},\mathbf{y}) \\
& 1 = \int f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y})
\end{aligned}
\tag{39}
$$

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

$$
\int f(\mathbf{x},\mathbf{y}) d\mathbf{x} \quad = \lambda U'(s(\mathbf{y})) \int f(\mathbf{x},\mathbf{y}) d\mathbf{x} + U'(s(\mathbf{y})) \int \mu(\mathbf{x},\mathbf{y}) d\mathbf{x} \tag{40}
$$

Noting that $\int f(\mathbf{x},\mathbf{y}) d\mathbf{x} = f(\mathbf{y})$ and rearranging gives the the following characterization of the optimal contract.

$$
\frac{1}{U'(s(\mathbf{y}))} \quad = \lambda + \left(\int \mu(\mathbf{x},\mathbf{y}) d\mathbf{x}\right) \frac{1}{f(\mathbf{y})}. \tag{41}
$$

Pointwise optimization of (39) with respect to f at $(\mathbf{x},\mathbf{y})$ yields the following closed form expression for $\mu(\mathbf{x},\mathbf{y})$.

$$
\mu(\mathbf{x},\mathbf{y}) \quad = \quad f(\mathbf{x},\mathbf{y}) \left(B(\mathbf{x}) - s(\mathbf{y}) - \eta\right) \tag{42}
$$

Substituting $\mu(\mathbf{x},\mathbf{y})$ into (41) gives:

$$
\begin{aligned}
& \frac{1}{U'(s(\mathbf{y}))} = \lambda + \left(\int \left(B(\mathbf{x}) - s(\mathbf{y}) - \eta\right) f(\mathbf{x},\mathbf{y}) d\mathbf{x}\right) \frac{1}{f(\mathbf{y})} \\
\iff \quad & \frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \int B(\mathbf{x}) \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} d\mathbf{x} - s(\mathbf{y}) \\
\iff \quad & \frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \int B(\mathbf{x}) f(\mathbf{x}|\mathbf{y}) d\mathbf{x} - s(\mathbf{y}).
\end{aligned}
\tag{43}
$$

Now we show that $f(\mathbf{x}|\mathbf{y}) = g(\mathbf{x}|\mathbf{y})$. First, observe that

$$
f(\mathbf{y}) \quad = \quad \int f(\mathbf{x},\mathbf{y}) dx \quad = \quad \int g(\mathbf{x},\mathbf{y}) e^{U(s(\mathbf{y})) - 1 - \nu} dx \quad = \quad g(\mathbf{y}) e^{U(s(\mathbf{y})) - 1 - \nu}, \tag{44}
$$

where the second equality follows from equation (18) and the first and third equalities are by definition of a marginal distribution. Equation (44) implies that

$$f(\mathbf{x}|\mathbf{y}) \;=\; \tfrac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} \;=\; \tfrac{g(\mathbf{x},\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}}{g(\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu}} \;=\; \tfrac{g(\mathbf{x},\mathbf{y})}{g(\mathbf{y})} \;=\; g(\mathbf{x}|\mathbf{y}). \tag{45}$$

Then the integral in the last line of (43) can be rewritten as

$$\textstyle\int B(\mathbf{x})f(\mathbf{x}|\mathbf{y})d\mathbf{x} \;=\; \int B(\mathbf{x})g(\mathbf{x}|\mathbf{y})d\mathbf{x} \;\equiv\; \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}]. \tag{46}$$

□

Proof of Proposition 3. Given a contract $s(\mathbf{y})$, the risk-neutral agent chooses $f(\mathbf{x},\mathbf{y})$ to maximize his expected utility minus his personal cost.

$$\begin{aligned} &\max_{f} \int s(\mathbf{y}) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) - \int f(\mathbf{x},\mathbf{y}) \ln(\tfrac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}) d(\mathbf{x},\mathbf{y}) \\ &\text{s.t.} \quad 1 = \int f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{47}$$

With ν as the multiplier on the constraint, pointwise optimization yields:

$$f(\mathbf{x},\mathbf{y}) \;\; = \;\; g(\mathbf{x},\mathbf{y}) e^{s(\mathbf{y})-\nu-1} \tag{48}$$

The principal's program is as follows, where we add and subtract ν on the left-hand side of the IR constraint.

$$\begin{aligned} \max_{s,f,\nu} \quad & \int (B(\mathbf{x}) - s(\mathbf{y})) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \nu + \int (s(\mathbf{y}) - \nu) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) - \int f(\mathbf{x},\mathbf{y}) \ln(\tfrac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}) d(\mathbf{x},\mathbf{y}) \geq \bar{U} \\ & f(\mathbf{x},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) e^{s(\mathbf{y})-\nu-1} \text{ for all } (\mathbf{x},\mathbf{y}) \\ & 1 = \int f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{49}$$

Substitute the IC constraints into the objective function and into the other constraints; this reduces the IR constraint to $\nu + 1 \geq \bar{U}$. Setting $\bar{U} = 0$ and binding the IR constraint yields $\nu + 1 = 0$. Substituting this into the objective and the other constraints, the principal's program can be rewritten as follows.

$$\begin{aligned} \max_{s} \int (B(\mathbf{x}) - s(\mathbf{y})) e^{s(\mathbf{y})} g(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad 1 = \int e^{s(\mathbf{y})} g(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{50}$$

Pointwise optimization (at $\mathbf{y}$) characterizes the optimal contract as follows.

$$\begin{aligned} e^{s(\mathbf{y})} \int B(\mathbf{x}) g(\mathbf{x},\mathbf{y}) d\mathbf{x} \;\; &= \;\; (s(\mathbf{y}) + 1) e^{s(\mathbf{y})} \int g(\mathbf{x},\mathbf{y}) d\mathbf{x} + \eta e^{s(\mathbf{y})} \int g(\mathbf{x},\mathbf{y}) d\mathbf{x} \\ \iff \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] \;\; &= \;\; s(\mathbf{y}) + 1 + \eta \end{aligned} \tag{51}$$

Then we can substitute $s(\mathbf{y}) = \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] - \eta - 1$ into equation (48) to obtain the equilibrium distribution,

$$f(\mathbf{x},\mathbf{y}) \;\; = \;\; g(\mathbf{x},\mathbf{y}) e^{\mathbb{E}_g[B(\mathbf{x})|\mathbf{y}]-\eta-1}. \tag{52}$$

By definition of a conditional multivariate normal distribution, $\mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] = \mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y}$.

Then we can rewrite the optimal contract and action as follows.

$$\begin{aligned} s(\mathbf{y}) &= \mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} - \eta - 1 \\ f(\mathbf{x}, \mathbf{y}) &= g(\mathbf{x}, \mathbf{y}) e^{\mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} - \eta - 1} \end{aligned} \tag{53}$$

Now letting $\hat{\mathbf{x}} = \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y}$, the solution above is equivalent to the following.

$$\begin{aligned} \hat{\mathbf{x}} &= \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} \\ \pi(\hat{\mathbf{x}}) &= \mathbf{b}^T \hat{\mathbf{x}} \\ s(\pi(\hat{\mathbf{x}})) &= B(\hat{\mathbf{x}}) - \eta - 1 \\ f(\mathbf{x}, \mathbf{y}) &= g(\mathbf{x}, \mathbf{y}) e^{\mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} - \eta - 1} \end{aligned} \tag{54}$$

Let $\mathbf{z} = [\mathbf{x}^T, \mathbf{y}^T]^T$ and $\mu = [\mu_{\mathbf{x}}^T, \mu_{\mathbf{y}}^T]^T$. Conjecture that $f(\mathbf{x}, \mathbf{y}) \sim \mathcal{N}(\mu, \Sigma)$, where $\Sigma = \Sigma_g \equiv \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix}$. Define $\Sigma^{-1} \equiv \begin{bmatrix} \tilde{\Sigma}_{\mathbf{xx}} & \tilde{\Sigma}_{\mathbf{xy}} \\ \tilde{\Sigma}_{\mathbf{yx}} & \tilde{\Sigma}_{\mathbf{yy}} \end{bmatrix}$. Then:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= (2\pi)^{-\frac{n+m}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(\mathbf{z} - \mu)^T \Sigma^{-1} (\mathbf{z} - \mu)) \\ &= g(\mathbf{x}, \mathbf{y}) \exp(\mathbf{z}^T \Sigma^{-1} \mu - \tfrac{1}{2} \mu^T \Sigma^{-1} \mu). \end{aligned} \tag{55}$$

It follows from (54) and (55) that the conjecture is true if there exists some μ satisfying the following conditions:

$$\begin{aligned} \mathbf{z}^T \Sigma^{-1} \mu - \tfrac{1}{2} \mu^T \Sigma^{-1} \mu &= \mathbf{y}^T \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} - \eta - 1 \\ \iff \mathbf{x}^T (\tilde{\Sigma}_{\mathbf{xx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{xy}} \mu_{\mathbf{y}}) + \mathbf{y}^T (\tilde{\Sigma}_{\mathbf{yx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{yy}} \mu_{\mathbf{y}}) - \tfrac{1}{2} \mu^T \Sigma^{-1} \mu &= \mathbf{y}^T \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} - \eta - 1 \end{aligned}$$

Since the coefficients on the variables and the constants must be equal on both sides, the expression gives the following three equalities.

$$\begin{aligned} \tilde{\Sigma}_{\mathbf{xx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{xy}} \mu_{\mathbf{y}} &= \mathbf{0}, \\ \tilde{\Sigma}_{\mathbf{yx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{yy}} \mu_{\mathbf{y}} &= \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b}, \\ \eta + 1 &= \tfrac{1}{2} \mu^T \Sigma^{-1} \mu. \end{aligned} \tag{56}$$

The first two expressions imply that $\Sigma^{-1} \mu = \begin{bmatrix} \mathbf{0} \\ \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} \end{bmatrix}$. Solving for μ:

$$\mu = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} \\ \Sigma_{\mathbf{yx}} \mathbf{b} \end{bmatrix} \tag{57}$$

Since $\Sigma_{\mathbf{xx}}, \Sigma_{\mathbf{yy}}$ are both invertible, we can solve for $\eta + 1$ as follows.

$$\begin{aligned}
\eta + 1 &= \tfrac{1}{2}\mu^T \Sigma^{-1} \mu \\
&= \tfrac{1}{2}\mu^T \begin{bmatrix} (\Sigma_{\mathbf{xx}} - \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}})^{-1} & \mathbf{0} \\ \mathbf{0} & (\Sigma_{\mathbf{yy}} - \Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1}\Sigma_{\mathbf{xy}})^{-1} \end{bmatrix} \begin{bmatrix} I & -\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1} \\ -\Sigma_{\mathbf{yx}}\Sigma_{\mathbf{xx}}^{-1} & I \end{bmatrix} \mu \\
&= \tfrac{1}{2}\mu^T \begin{bmatrix} \mathbf{0} \\ \Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{b} \end{bmatrix} = \tfrac{1}{2}\mathbf{b}^T \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\Sigma_{\mathbf{yx}}\mathbf{b}
\end{aligned}$$

□

References

Amershi, A. H. and P. Cheng. 'On the demand for historical events recording and maintenance of audit trails.' Contemporary Accounting Research 6.1 (1989): 72–90.

Arya, A. and J. C. Glover. 'On the upsides of aggregation.' Journal of Management Accounting Research 26.2 (2014): 151–166.

Banker, R. D. and S. M. Datar. 'Sensitivity, precision, and linear aggregation of signals for performance evaluation.' Journal of Accounting Research 27.1 (1989): 21–39.

Bertomeu, J. 'Risk management and the design of efficient incentive schemes.' Diss. Carnegie Mellon University, 2008.

Bloomfield, M., et al. 'Cost shielding in executive bonus plans.' Journal of Accounting and Economics (2021). In press.

Bonham, J. 'Shaping Incentives through Measurement and Contracts.' Unpublished paper, 2020. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3691705.

Bonham, J. and A. Riggs-Cragun. 'Contracting on What Firm Owners Value.' Unpublished paper, University of Chicago Booth School of Business, 2021. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3892838.

Burgstahler, D. and I. Dichev. 'Earnings management to avoid earnings decreases and losses.' Journal of Accounting and Economics 24.1 (1997): 99–126.

Bushman, R. M. and A. J. Smith. 'Financial accounting information and corporate governance.' Journal of Accounting and Economics 32.1-3 (2001): 237–333.

Core, J. E., W. R. Guay and R. E. Verrecchia. 'Price versus non-price performance measures in optimal CEO compensation contracts.' The Accounting Review 78.4 (2003): 957–981.

Cover, T. M. and J. A. Thomas. Elements of Information Theory, 2nd Edition. John Wiley & Sons, Inc., 2006.

De Angelis, D. and Y. Grinstein. 'Performance terms in CEO compensation contracts.' Review of Finance 19.2 (2015): 619–651.

FASB. Conceptual Framework for Financial Reporting: Objective of Financial Reporting and Qualitative Decision-Useful Financial Reporting Information. Norwalk, CT: Financial Accounting Standards Board, 2006.

Feltham, G. A. and J. Xie. 'Performance measure congruity and diversity in multi-task principal/agent relations.' The Accounting Review 69.3 (1994): 429–453.

Gao, P. 'A measurement approach to conservatism and earnings management.' Journal of Accounting and Economics 55.2-3 (2013): 251–268.

Hart, O. and B. Holmström. 'The theory of contracts.' Advances in Economic Theory: Fifth World Congress. Cambridge University Press Cambridge. 1987.

Hébert, B. 'Moral hazard and the optimality of debt.' The Review of Economic Studies 85.4 (2018): 2214–2252.

Hellwig, M. F. 'The role of boundary solutions in principal–agent problems of the Holmström–Milgrom type.' Journal of Economic Theory 136.1 (2007): 446–475.

Hemmer, T. 'Optimal Dynamic Relative Performance Evaluation.' Unpublished paper, Rice University Jesse H. Jones Graduate School of Business, 2017. Available at https://papers.ssrn.com/sol3/papers.cfm?abstract_id=2972313.

Holmström, B. 'Moral hazard and observability.' The Bell Journal of Economics 10.1 (1979): 74–91.

Holmström, B. and P. Milgrom. 'Aggregation and linearity in the provision of intertemporal incentives.' Econometrica 55.2 (1987): 303–328.

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

Lambert, R. A. 'Contracting theory and accounting.' Journal of Accounting and Economics 32.1-3 (2001): 3–87.

Leuz, C. 'Embedded contingencies in earnings-based contracts: A note on the role of accruals in contracting.' Unpublished paper, 1999.

Meth, B. 'Reduction of outcome variance: Optimality and incentives.' Contemporary Accounting Research 13.1 (1996): 309–328.

Roychowdhury, S. 'Earnings management through real activities manipulation.' Journal of accounting and economics 42.3 (2006): 335–370.

Watts, R. L. 'Conservatism in accounting part I: Explanations and implications.' Accounting Horizons 17.3 (2003): 207–221.