Contracting on Aggregated Accounting Estimates

Jonathan Bonham
University of Chicago
Booth School of Business

Amoray Riggs-Cragun
University of Chicago
Booth School of Business

November 9, 2022

Abstract

Using a principal-agent setting in which the agent has a rich action space, we provide a theoretical justification for contracting on highly-aggregated accounting metrics. We show that the optimal contracting process can be decomposed into three stages: 1) constructing unbiased estimates of items that the principal values, 2) aggregating those estimates using the weights in the principal's objective (as opposed to weighting by sensitivity or precision), and 3) compensating the agent on the aggregated estimate. This process mirrors how accounting measurement, aggregation and executive compensation are done in practice. Our results reconcile the conflict between the stewardship and valuation uses of information; when the agent has flexible control over firm performance, evaluating the manager and valuing the firm are one and the same. In a tractable specification of our model in which normal distributions arise endogenously, we show that optimal measurement rules are conservative yet produce unbiased estimates. Moreover, a weaker link between investment and future returns warrants more conservative treatment of expected future benefits, providing a rationale for the immediate expensing of R&D, the capitalization of PP&E and the accrual of credit sales.

Keywords: Optimal contracting, accounting, aggregation, measurement.

JEL Classification: D86, J41

We appreciate helpful comments from Jeremy Bertomeu (discussant), Robert Göx, Christian Laux, Christian Leuz, Valeri Nikolaev, Haresh Sapra, Abbie Smith, Frank Zhou and participants in the DAR & DART Accounting Theory Seminar, the Chicago Booth Accounting Research Workshop, the Accounting and Economics Society webinar, Columbia Business School, the Wharton School of the University of Pennsylvania, and the EIASM Workshop on Accounting and Economics. 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 committees have access to enormous databases of information for evaluating and compensating executives, but executive pay is typically tied to a few highly aggregated accounting metrics such as total revenue or net income.[1] In practice, this reduction of rich data into a few metrics occurs on two levels. The first, which we call *measurement*, is the process by which observable transactions and events are used to estimate the individual accounting items recorded in journal entries. The second, which we call *aggregation*, is the linear process by which individual entries are summed up to form aggregated metrics, such as those that appear on the financial statements.

Classical agency theory suggests that these aggregated accounting metrics are inefficient for contracting. For example, the results from Banker and Datar [1989] indicate that accounts should be weighted according to their relative sensitivity and precision, rather than being added up with weights of 1 or -1 according to double-entry conventions. More generally, prior theory shows a conflict between the valuation and contracting uses of information,[2] raising the question of why metrics generated by the measurement and aggregation rules of U.S. GAAP – a valuation-oriented system – would be used for contracting purposes. Absent imposing additional frictions on the moral hazard problem (such as costs of contracting complexity), agency theory has yet to explain why executive compensation contracts are conditioned on these highly-aggregated, valuation-oriented accounting metrics.

Using an unconventional approach to moral hazard where the agent has a very rich action space, we provide a framework for measurement and aggregation that closely maps to how accounting is done in practice. Our framework rectifies the conflict between the valuation and contracting uses of information and shows that contracting on aggregated accounting metrics can be just as efficient as contracting on all of the data used to measure their underlying components.

We draw our conclusions from an agency model in which a principal cares about a set of unobservable constructs, $\mathbf{x}$, which she values according to the linear function $B(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$, which we term the principal's *objective*.[3] We assume that the constructs $\mathbf{x}$ are not directly

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

[2]See Gjesdal [1981], Paul [1992], Feltham and Xie [1994], and Lambert [2001].

[3]For example, $B(\mathbf{x})$ could be net present value, where $\mathbf{x}$ are cash flows that arrive in different

contractible, but there is a vector of observable data $\mathbf{y}$ available to the principal. To form the basis of a legally-enforceable contract, $\mathbf{y}$ must be objective and verifiable; building on prior ideas from accounting thought, we interpret $\mathbf{y}$ as the set of verifiable exchanges, transaction characteristics, and events relevant to the contracting period.[4]

We assume that the agent chooses a joint distribution, $f(\mathbf{x}, \mathbf{y})$, directly from the space of all possible distributions over $\mathbf{x}$ and $\mathbf{y}$. This is the key assumption that drives our results and diverges from prior work. Beginning with the seminal work of Holmström [1979], agency theory has been dominated by the *parameterized distribution formulation* of the moral hazard problem, which models the agent as affecting the parameter(s) of a distribution whose functional form is outside of the agent's control. We take an alternative (though not novel) approach, adopting 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 rich action space captures the immense flexibility that executives have in running a company. Executives might affect variance by adjusting the riskiness of the firm's project portfolio, increase skewness by allocating investments to exploratory R&D projects, and introduce discontinuities by managing earnings around salient thresholds.

Solving the principal's program, we first characterize the optimal contract conditioned on all of the data in $\mathbf{y}$. This is akin to a contract conditioned on every piece of verifiable data relevant to the firm, which is not descriptive of practice. For public firms, such a contract would be practically infeasible due to the sheer volume of transactional data (for example, Amazon ships 584 million packages a year). We therefore turn to optimal data reduction – a central function of accounting – and examine composite metrics that are *efficient* or *(weakly) optimal* for contracting purposes, meaning that there is no loss to contracting on the composite metric(s) relative to contracting on all of the data in $\mathbf{y}$.

periods and $\mathbf{b}$ is a vector of discount factors. Alternatively, $B(\mathbf{x})$ might represent Hicksian income, where $\mathbf{x}$ consists of increases and decreases in shareholder wealth and $\mathbf{b}$ consists of corresponding weights of one or negative one.

[4]Butterworth, Gibbins, and King [1982] suggest that the set of exchange transactions between the firm and outside parties provide a natural basis for contracting. Likewise, Paton and Littleton [1940] note that the inflowing and outflowing "facts of services rendered" are generally objective, and Leuz [1998] suggests that past transactions and events are verifiable and are likely to contain information pertinent to prevailing incentive problems.

We show that the data $\mathbf{y}$ enters the contract through the conditional expectation

$$\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}] = \mathbf{b}^T\,\mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}], \tag{1}$$

where $\mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}]$ are the expected values of the constructs $\mathbf{x}$ given the data $\mathbf{y}$ under the equilibrium action f. Hence, a contract conditioned on the single aggregate $B(\hat{\mathbf{x}}) \equiv \mathbf{b}^T\,\mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}]$ is just as efficient as the highly complex contract conditioned on all of the data in $\mathbf{y}$.

This suggests a very natural decomposition of the optimal contracting process into three distinct stages: measurement, aggregation, and compensation. In the *measurement* stage, the principal uses observable data $\mathbf{y}$ to construct *estimates*, $\hat{\mathbf{x}} \equiv \mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}]$, of the unobservable constructs $\mathbf{x}$. This is akin to using observable transaction characteristics and events ($\mathbf{y}$) – such as delivery status, cash collections, payments to suppliers, and customer credit history – to produce imperfect measures ($\hat{\mathbf{x}}$) of the FASB-defined constructs of revenues and expenses ($\mathbf{x}$). In the *aggregation* stage, the individual estimates $\hat{\mathbf{x}}$ are linearly aggregated according to the weights $\mathbf{b}$ to form one or more composite measures. For example, if the principal cares equally about the revenues from each product, these revenues are simply summed up into total revenues.[5] Finally, the principal conditions the agent's compensation contract on the aggregated estimate(s).

Our three-step decomposition can rationalize contracting on aggregated accounting metrics in practice. As noted by Butterworth, Gibbins, and King [1982], "Relative shares cannot be based on unobservable events," (that is, the principal cannot write a contract on the constructs $\mathbf{x}$), which creates the need for "some tentative measure of enterprise performance. . . for without it there can be no basis for assessment of the relative shares to which the holders of relative interests are entitled, and therefore no basis for a contract between them." Accounting measurement provides these "tentative measures," $\hat{\mathbf{x}}$, proxies for $\mathbf{x}$ that are contractible because they are based on verifiable information $\mathbf{y}$.[6] Our results show that

[5]Note that this simple "summing up" occurs in the *aggregation* stage, not the measurement stage. The measurement stage may generally be non-linear, as there are many contingencies built into the measurement process. Thus, if the principal cares equally about revenues from different products, then there is equally-weighted aggregation of the individual estimated product revenues $\hat{\mathbf{x}}$, but not of the data $\mathbf{y}$ used to estimate those revenues.

[6]As noted by Ijiri [1975], "income *per se* and an income figure obtained as a result of measuring income are two entirely different things" (p. 54). In our model, economic constructs such as "income *per se*" are captured by $\mathbf{x}$, and measures of those constructs are captured

accounting proxies $\hat{\mathbf{x}}$ are contractually efficient if they are unbiased ($\mathbb{E}[\hat{\mathbf{x}}] = \mathbb{E}[\mathbf{x}]$), or said differently, if they *faithfully represent* the underlying economic constructs they purport to represent, $\mathbf{x}$. In practice, the individual estimates $\hat{\mathbf{x}}$ are entered into the accounting system and linearly aggregated with a weight of 1 or -1 according to double entry conventions (e.g. credits or debits on the income statement), while in our model, the estimates are aggregated according to $\mathbf{b}$, the linear weights in the principal's objective. Therefore, our results rationalize contracting on a given aggregated accounting metric to the extent that 1) its components faithfully represent the economic constructs they purport to represent, and 2) the principal values the constructs equally, using the same debit/credit aggregation as a conventional accounting system.

Prior accounting theory has tackled aggregation and measurement separately; we will discuss our contribution to each in turn, beginning with aggregation. The standing authority on optimal aggregation is Banker and Datar [1989]. In their seminal paper, Banker and Datar show for certain classes of distributions, contractually 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. (Indeed, no reference to sensitivity or precision is made in the FASB definitions of financial accounting constructs.) 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 aggregated income statement items 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 those aggregated accounting metrics are inefficient for contracting.

Our framework produces very different results. Notice from (1) that the optimal contract can be conditioned on $B(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$. The optimal linear weights for contracting are given simply by $\mathbf{b}$; that is, the optimal weight on each variable is determined solely by how much the principal values it. For example, if shareholders intrinsically value each revenue and expense item equally – which is almost a truism to the extent that revenues and expenses are defined as changes in shareholder wealth – then an optimal compensation contract can

by $\hat{\mathbf{x}}$. Ijiri also notes that "the measurement process must begin with verifiable facts" (p. 36); these verifiable facts are captured in our model by $\mathbf{y}$.

be written on net income, which is an equally-weighted aggregate of all revenues and expenses. Alternatively, if shareholders place a higher value on certain income statement items (because, for example, GAAP revenue and expense constructs do not exactly conform to changes in shareholder wealth), those items will be weighted more heavily for compensation purposes. Hence, in our setting there is no conflict between the valuation and compensation uses of information, consistent with the empirical findings of Bushman, Engel, and Smith [2006] and Banker, Huang, and Natarajan [2009].

The intuition for this result stems from the agent's rich action space under the generalized distribution approach, which creates a one-to-one mapping between the contract offered and the distribution implemented. Whereas in the classic parametric approach the principal can choose the best possible contract that implements a given distribution, under the generalized approach, any change in the contract causes the agent to change his action. If the principal attempts to squeeze risk-sharing efficiency from the contract by, for example, weighting the signals by their relative precision, the agent will respond by taking an action that maximizes the precision-weighted aggregate rather than the principal's objective. Therefore, the best the principal can do is hand control of her objective over to the agent (consistent with a complete separation of ownership and control, e.g. Fama and Jensen [1983]).

Our aggregation results may have useful implications for 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."

Our paper also contributes to the literature on optimal measurement. Accounting theory has typically modeled measurement as a biased, one-step mapping from underlying fundamentals (e.g., $\mathbf{x}$) to performance measures (e.g., $\hat{\mathbf{x}}$). As in Gao [2013], our framework takes a two-step approach to accounting measurement. The first step is from unobservable firm fundamentals, $\mathbf{x}$, to observable transaction characteristics, $\mathbf{y}$. This mapping is dictated by $f(\mathbf{y}|\mathbf{x})$, which is under the control of the agent by his choice of $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x})f(\mathbf{y}|\mathbf{x})$. One way to interpret this step is that while the principal cannot directly observe the underlying constructs she values ($\mathbf{x}$) or the agent's actions (f), the agent's actions leave "traces" in the form of $\mathbf{y}$, observable and verifiable data such as transaction characteristics. Alternatively, as in Gao [2013] the agent may take actions that distort transaction characteristics without improving the underlying constructs for which they provide evidence; for example, the agent might engage in channel stuffing to inflate inventory deliveries without improving sales. The second step is from the transaction characteristics $\mathbf{y}$ to the estimates $\hat{\mathbf{x}}$, which in equilibrium are unbiased conditional expectations $\hat{\mathbf{x}} = \mathbb{E}_f[\mathbf{x} \,|\, \mathbf{y}]$. This step is under the control of the principal, who writes measurement rules $\hat{\mathbf{x}}(\mathbf{y})$ as part of the solution to her contracting problem.

An *optimal measurement rule* $\hat{\mathbf{x}}(\mathbf{y})$ dictates *how* observable transaction characteristics $\mathbf{y}$ should be used to construct an unbiased estimate of $\mathbf{x}$. To study optimal measurement, 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.

The extent of conservative bias required in the measurement stage depends on the ease with which managers can game transaction characteristics. This has important implications when accounting for investments with uncertain future returns, a central measurement issue in accounting. To illustrate these implications, we present a simple application of our model which suggests that optimal measurement depends on the availability of reliable evidence about future returns. We find that when manipulating evidence about future returns is infinitely costly to the manager, fair value accounting is optimal, but in all other cases, the optimal measurement rule for a given investment *depends on the likelihood that the investment will produce future returns.* In particular, the optimal measure is increasing in the correlation between investment and future returns. When the correlation between an investment and future returns is zero, the optimal measurement rule reports the cash outlay and makes no attempt to estimate future returns; that is, the investment is fully expensed. As the correlation between investment and future returns increases, the optimal measurement rule expenses some, but not all, of the investment outlay. As the correlation increases further, the optimal measurement rule begins to incorporate some future returns, and when the correlation is perfect, all future returns are included in income.

Although we abstract away from accrual reversals, the optimal measurement rule described above applies to many conventions in accrual accounting. In cases where future returns from an investment are highly uncertain, such as with R&D or advertising, GAAP generally prescribes immediate expensing. In a credit sale, the sacrifice of inventory (the "investment") is highly correlated with the collection of cash; in this case, GAAP prescribes revenue accrual. Investments in PP&E arguably have a moderate correlation with future return; in this case, GAAP dictates deducting some, but not all, of the capital expenditure as depreciation expense, but does *not* allow future returns from the investment to be accrued.

The pattern across these GAAP practices is that measurement of an uncertain investment depends on the likelihood that the investment will pay off, which is exactly what we find. This idea was also expressed by Ordelheide [1988], who (roughly translated) notes that whether a cash outlay is recorded as an asset depends on whether the cash outlay is an indicator of future cash receipts. More generally, our optimal measurement rule for uncertain investments is consistent with a basic feature of accrual accounting, that virtually every accrual contains an implicit assumption about future events (Beaver [1991]). Assets are *probable* future economic benefits, and hence, cost deferral and revenue accrual depends on the probability of the cost being recovered or the revenue being realized.

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] in a benchmark setting where measurement is perfect (i.e. $\mathbf{x}$ is observable). Here, 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 layer in the measurement process by assuming that the constructs $\mathbf{x}$ are unobservable; here, we provide a three-step decomposition of the contract that provides a theoretic rationale for contracting on aggregated accounting estimates like net income. In section 5, we develop a tractable specification of our model to investigate how the estimates of $\mathbf{x}$ are formed. We provide stylized applications that illustrate the contracting rationale for conservatism (section 5.1) and for conventional standards regarding accounting for uncertain investments (section 5.2). Section 6 concludes and provides suggestions for future work. Proofs are in the appendix.

2 Related Work

We study optimal data reduction on two levels, measurement and aggregation, that have been studied separately in prior research. The aggregation component of our paper addresses the same question tackled by Banker and Datar [1989]: 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 aggregated metric 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; intuitively, they show that aggregation is optimal when the cost of contracting on many variables is outweighed by the cost of the information loss resulting from aggregation. 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 introduced 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 constructs the principal cares about are unobservable, the optimal contract can be separated into three stages – measurement, aggregation and contracting. As with aggregation, the measurement stage may entail data reduction.[7] 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

[7]For example, assume that the principal cares only about one construct, x, such that $B(\mathbf{x}) = x$, and assume that x is unobservable but that the principal can observe the data $\mathbf{y}$. A (weakly) optimal measurement rule, $\hat{x}(\mathbf{y})$, uses the observable data $\mathbf{y}$ to produce a single estimate, $\hat{x} = \mathbb{E}_f[x|\,\mathbf{y}]$, such that there is no loss to contracting on the estimate relative to contracting on the underlying data.

same measures in multiple contracts.

We show that conditioning the contract on the aggregated estimate $B(\hat{\mathbf{x}}) \equiv \mathbf{b}^T\hat{\mathbf{x}}$ is just as efficient as conditioning on all of the data in $\mathbf{y}$. This rectifies the seemingly disjointed *valuation* versus *stewardship* uses of accounting information: the contractually optimal way to evaluate the manager using the data in $\mathbf{y}$ is to estimate the value of the principal's objective, $B(\mathbf{x})$. This result stems from the agent's flexible action space in our setting, suggesting that when CEOs have very flexible control over firm performance, valuing the firm and evaluating the manager are one and the same. Our results are in stark contrast to findings from the classic approach, which have predicted that information is used differently for valuation and stewardship (Gjesdal [1981], Paul [1992], Feltham and Xie [1994], Lambert [2001]).

The main difference between our paper and conventional agency theory papers like Banker and Datar [1989] is in how the agent's action is modeled. The bulk of the agency literature uses the *parameterized distribution formulation* of the moral hazard problem, which models the agent's action (often interpreted as "effort") as the choice of one or more parameters in the distribution(s) over relevant variables.[8] For example, if we were to use the parametric approach in our setting, we might model the agent as choosing an effort level, a, that parameterizes the distribution over $\mathbf{x}$ and $\mathbf{y}$, denoted $f(\mathbf{x}, \mathbf{y}; a)$. Notice that under this approach, the agent's choice of a entails choosing a distribution from a restricted parametric set; there is nothing he can do to break free of the functional form of $f(\mathbf{x}, \mathbf{y}; a)$.

In this paper, we depart from the conventional parametric approach and instead adopt the *generalized distribution formulation*; under this approach, the agent can directly choose *any* distribution $f(\mathbf{x}, \mathbf{y})$. This approach captures the rich opportunity set available to firm executives, with significant influence over the company's operations, product offerings, investment portfolio and competitive strategy. The sheer number of books proffering business advice (over 80,000 titles on Amazon) suggests that a CEO's inputs are far more complex

[8]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]); there, 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]). Notice that while the agent in this case can affect both mean and variance in this case, his influence over the distribution is still quite constrained. He has no way of affecting skewness or kurtosis or of introducing discontinuities; by assumption, the *only* distribution possible is a normal.

than simply exerting effort (else those books could be replaced by the maxim, "Try hard"). Moreover, 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]). Thus, the generalized approach seems particularly descriptive of the role played by corporate executives.[9]

The terms "parameterized distribution formulation" and "generalized distribution formulation" 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.

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.

The first paper to use the generalized approach was Holmström and Milgrom [1987]; in section 2 of their paper, they model the agent as directly choosing the probability of every

[9]Outside the C-suite, there may be situations where the classic parametric approach is more descriptive. A production line worker who functions as a "cog in the machine" is likely powerless to change the shape of the distribution; here, the parametric approach seems appropriate.

(discrete) outcome in a single-period model. Holmström and Milgrom [1987] are better known for what they do next. They divide the single period into subperiods and show that as the subperiod length approaches zero, the solution approximates a continuous time model in which the agent controls the drift of a Brownian motion, and moreover, that the optimal contract is linear in the ending position of the Brownian process. This approximation spawned the influential "LEN" model, which exogenously restricts contracts to be *L*inear, assumes that the agent has negative *E*xponential utility, and models the agent's action as affecting the mean of a *N*ormal distribution. Notice that the LEN modeling assumptions fall squarely under the parametric approach, because the agent's choice of distributions is restricted to a set of normal distributions with exogenously fixed variance. Interestingly, the tractable specification we develop in section 5 produces *solutions* that share several features with the LEN *assumptions* while remaining in a simple static setting; we do not introduce multiple periods or Brownian motions to achieve linear contracts or normally distributed performance measures.

Since Holmström and Milgrom [1987], very few papers have used the generalized distribution formulation, perhaps because (as indicated in the Hart and Holmström [1987] quote above) it is difficult to interpret the agent's personal cost of "choosing a distribution." An important recent development is Hébert [2018], who studies optimal security design using the generalized distribution formulation. Hébert [2018] 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, as do Bonham and Riggs-Cragun [2021], a close predecessor to this paper.

Bonham and Riggs-Cragun [2021] 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 Bonham and Riggs-Cragun [2021] assume that the principal cares about a single variable, x, we assume that the principal cares about *many* variables, $\mathbf{x}$. This allows us to study optimal aggregation, an important feature of accounting. Second, 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, a few other papers 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 [2021] uses the generalized approach to study how measurement and contracts shape productive incentives. In Bonham [2021], 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, 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 Benchmark model: Contracting on aggregates

A principal cares about a set of unobservable constructs, $\mathbf{x} \equiv (x_1 \dots x_m)^T$. She might value some constructs differently than others; we say that the principal values $\mathbf{x} \in \mathbb{R}^m$ according to her *objective*, $B(\mathbf{x})$, where $B : \mathbb{R}^m \to \mathbb{R}$. We will typically assume that the principal values the constructs linearly such that $B(\mathbf{x}) = \mathbf{b}^T \mathbf{x}$. For example, $B(\mathbf{x})$ might be the net present value of the firm, where $\mathbf{x}$ are cash flows that arrive in different periods and $\mathbf{b}$ is a vector of discount factors. Or $B(\mathbf{x})$ might be Hicksian income,[10] where $\mathbf{x}$ consists of increases and decreases in shareholder wealth (i.e. net economic assets). In this case, $b_i = 1$ for $x_i > 0$ and $b_j = -1$ for $x_j < 0$; without loss of generality, we can group the elements of $\mathbf{x}$ such that $x_1 \dots x_t$ are increases in shareholder wealth and $x_{t+1} \dots x_m$ are decreases in shareholder wealth, such that the principal's objective of Hicksian income can be expressed as follows.

$$B(\mathbf{x}) = \underbrace{x_1 + x_2 + \ldots + x_t}_{\substack{\text{increases in} \\ \text{wealth}}} - \underbrace{(x_{t+1} + \ldots + x_m)}_{\substack{\text{decreases in} \\ \text{wealth}}} \tag{2}$$

Continuing with the example wherein $B(\mathbf{x})$ is Hicksian income, $\mathbf{x}$ could be defined as "earned" changes in wealth according to the FASB definitions of revenues and expenses.[11]

[10] *Hicksian income* is the change in the firm's net economic assets other than from transactions with owners, or equivalently, the amount that can be paid out in dividends during a period while leaving the firm as well off at the end of the period as it was at the beginning of the period (Hicks [1939]).

[11] FASB defines *revenues* as inflows or other enhancements of assets of an entity or settlements

If FASB-defined revenues and expenses perfectly reflected changes in shareholder wealth, the principal's objective of Hicksian income would still be given by (2). But if the FASB definitions do not align with changes in shareholder wealth, then $\mathbf{b}$ might be a set of weighting adjustments to GAAP revenues and expenses such that the aggregate $B(\mathbf{x})$ is Hicksian income. For example, assume that the FASB constructs are defined by changes in shareholder wealth with the exception of one revenue item, x_1, which understates the true change in shareholder wealth due to a stringent revenue recognition criterion. Then $\mathbf{b}$ consists of weights of -1 on all expenses and 1 on all revenues, with the exception of $b_1 > 1$, which adjusts the revenue item to reflect the true increase in shareholder wealth such that $B(\mathbf{x})$ reflects Hicksian income.

$$B(\mathbf{x}) = \underbrace{b_1 x_1 + x_2 + x_3 + \ldots + x_t}_{\text{adjusted GAAP revenues}} - \underbrace{(x_{t+1} + \ldots + x_m)}_{\text{GAAP expenses}} \tag{3}$$

These examples illustrate that the weights $\mathbf{b}$ in the principal's objective depend to an extent on how the constructs $\mathbf{x}$ are defined. A particular interpretation of the constructs $\mathbf{x}$ is not needed for our analysis, but to contextualize our results we will often refer back to the example of $\mathbf{x}$ as FASB-defined constructs of revenues and expenses.

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, the principal designs a contract, s. We assume in this section that the constructs $\mathbf{x}$ are directly contractible, i.e. the principal writes a contract, $s(\mathbf{x})$, where $s : \mathbb{R}^m \to \mathbb{R}$ pays the agent $s(\mathbf{x})$ when the outcome $\mathbf{x}$ is realized. This benchmark assumption allows us to study optimal aggregation in isolation and facilitates a straightforward comparison to Banker and Datar [1989]. Let $U(s)$ denote the utility from compensation of a (weakly) risk averse agent, 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 $\bar{U}$. Finally, assume that the principal is risk neutral such that her utility is equal to her net payoff, $B(\mathbf{x}) - s(\cdot)$.

We study two questions in this section. First, under what conditions can the optimal contract be conditioned on a linear aggregate of the outcomes $x_1, \ldots, x_m$ such that there is

of its liabilities (or a combination of both) from delivering or producing goods, rendering services, or other activities that constitute the entity's ongoing major or central operations. FASB defines *expenses* as outflows or other using up of assets or incurrences of liabilities (or a combination of both) from delivering or producing goods, rendering services, or carrying out other activities that constitute the entity's ongoing major or central operations.

no loss relative to conditioning the contract on all m outcomes? Second, in cases where such an optimal aggregate exists, 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 conventional parameterized distribution 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 effort level $a \in A \subseteq \mathbb{R}$, where a parameterizes 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 accepts the contract. 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, where $f_a(\mathbf{x}; a)$ is the derivative of $f(\mathbf{x}; a)$ with respect to a and $\int d\,\mathbf{x}$ denotes integration over $x_1, x_2, \ldots, x_m$.

$$\begin{aligned} \max_{s,a} \quad & \int (B(\mathbf{x}) - s(\mathbf{x})) f(\mathbf{x}; a) d\mathbf{x} \\ \text{s.t.} \quad & \int \left(U(s(\mathbf{x})) - V(a)\right) f(\mathbf{x}; a) d\,\mathbf{x} \geq \bar{U} \\ \text{and} \quad & \int U(s(\mathbf{x})) f_a(\mathbf{x}; a) d\mathbf{x} = V'(a) \end{aligned} \tag{4}$$

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.

$$\frac{1}{U'(s(\mathbf{x}))} = \lambda + \mu \cdot \frac{f_a(\mathbf{x}; a)}{f(\mathbf{x}; a)} \tag{5}$$

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 (5) for $s(\mathbf{x})$.

$$s(\mathbf{x}) \quad = \quad U'^{-1}\left(1/\left(\lambda + \mu \cdot \frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)}\right)\right). \tag{6}$$

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 contract that is conditioned on that measure.

$$\begin{array}{rrcl} \textit{Aggregator:} & \pi(\mathbf{x}) & \equiv & \lambda + \mu \cdot \frac{f_a(\mathbf{x};a)}{f(\mathbf{x};a)} \\ \textit{Contract:} & s(\pi(\mathbf{x})) & = & U'^{-1}\left(\frac{1}{\pi(\mathbf{x})}\right) \end{array} \tag{7}$$

This decomposition shows that $\pi(\mathbf{x})$ is an optimal linear aggregator for contracting; that is, there is no loss to conditioning the contract on $\pi(\mathbf{x})$ relative to conditioning the contract on every element of $\mathbf{x}$. Banker and Datar [1989] observe 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.[12] They then characterize 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.

Banker and Datar [1989] are best known for what they do next. Within a subclass of distributions for which some linear aggregator is optimal, they characterize the relative linear weights on the measures by their signal-to-noise ratios. Letting $\pi(\mathbf{x}) = \sum_j \pi_j x_j$, the relative weights on measures x_i and x_j in this optimal linear aggregator are characterized by:

$$\frac{\pi_i}{\pi_j} \quad = \quad \frac{\partial E(x_i|a)/\partial a}{\mathrm{Var}(x_i)} \Big/ \frac{\partial E(x_j|a)/\partial a}{\mathrm{Var}(x_j)} \tag{8}$$

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 this result why executive compensation contracts are routinely conditioned

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

on aggregated accounting totals rather than on their (readily available and contractible) disaggregated components.

3.2 The generalized distribution approach

We now revisit optimal linear aggregation by invoking the *generalized distribution approach.* Assume that the agent chooses a distribution $f(\mathbf{x}) \in \Delta(\mathbf{x})$, where $\Delta(\mathbf{x})$ denotes the space of all probability distributions over $\mathbf{x}$. 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.

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. Under the classic approach, 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 relaxes the parametric restriction and assumes that the agent can choose *any* distribution; that is, the agent selects $f(\mathbf{x})$ from $\Delta(\mathbf{x}) \equiv \{f(\mathbf{x}) | \int f(\mathbf{x}) d\mathbf{x} = 1, f(\mathbf{x}) \geq 0 \ \forall \mathbf{x}\}$. 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 distribution 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 $D(f(\mathbf{x}) || g(\mathbf{x}))$.

$$V(f) \quad = \quad D(f(\mathbf{x}) || g(\mathbf{x})) \quad \equiv \quad \int f(\mathbf{x}) \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x}. \tag{9}$$

Divergence (or *relative entropy*) measures the dissimilarity between two distributions and has many applications in information theory and machine learning. For our purposes, it captures the personal cost incurred by the agent when he takes the requisite actions to implement 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.[13]

[13]Note that this cost function cannot capture *costless* changes in risk. This rules out settings

The KL-divergence cost function has several appealing properties that are discussed in Bonham and Riggs-Cragun [2021] and repeated here for reference. 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 $\mathbf{x}$. 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. Finally, KL divergence is generally asymmetric ($D(f||g) \neq D(g||f)$). This reflects the idea that the personal cost of taking an action tends to differ from the cost of undoing an action; for example, the cost of launching a product and discontinuing a product need not be the same.

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, she 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. The distribution g is undesirable from the principal's perspective but is maximally desirable from the agent's perspective, absent contractual incentives.

The distribution g warrants further discussion. This exogenously specified distribution is the distribution that the agent would implement if offered a flat wage; that is, it is the distribution that arises when the agent does *exactly what he wants*, absent contractual incentives. We are agnostic about these preferences. For an extremely effort-averse CEO, g might be the distribution arising when the agent spends all his time watching Netflix instead of working. For a CEO seeking the "quiet life," g might be the distribution that keeps

such as a CEO using derivatives to transfer cash flows across states *at no personal cost* (but it does not rule out cases where the CEO has to exert effort to find the right combination of derivatives to perfectly hedge against risk).

the company humming along modestly, such that the CEO does not have to work too hard but his reputation is not destroyed. For a CEO who likes empire building, g might be a distribution that rapidly grows the company at the expense of long-term value.

Notice that because g is by definition the agent's cost-minimizing distribution, *any* divergence from g is costly, including in the downward direction. Hence, the agent incurs disutility by actively working to destroy output, beyond any output destruction inherent in g. There could be settings in which costly downward divergences are descriptive. For example, g may represent the distribution wherein the CEO works just hard enough that his reputation is not damaged; here, a downward divergence from g is personally costly not because it takes "effort," but because abnormally poor performance hurts the CEO's future career prospects. Regardless of whether costly downward divergences are descriptive, they are never incentivized in equilibrium and hence have no impact on our results.

With all of our modeling assumptions in place, we turn to the agent's problem. When offered contract s, the agent chooses f to maximize his expected utility from compensation less his personal cost.

$$\begin{aligned} \max_{f} \quad & \int U(s(\mathbf{x}))f(\mathbf{x})d\mathbf{x} - \int f(\mathbf{x}) \ln\left(\tfrac{f(\mathbf{x})}{g(\mathbf{x})}\right) d\mathbf{x} \\ \text{s.t.} \quad & 1 = \int f(\mathbf{x})d\mathbf{x} \\ & f(\mathbf{x}) \geq 0 \text{ for all } \mathbf{x} \end{aligned} \tag{10}$$

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 (10) yields the following incentive compatible action.

$$f(\mathbf{x}) \quad = \quad g(\mathbf{x})e^{U(s(\mathbf{x}))-\nu-1}, \tag{11}$$

where $\nu = \ln\left(\int g(\mathbf{x})e^{U(s(\mathbf{x}))-1}d\mathbf{x}\right)$ is obtained by substituting (11) into the constraint $1 = \int f(\mathbf{x})d\mathbf{x}$. 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 (11) is maintained for all $\mathbf{x}$. Taking into account that the agent will respond in this way, the principal solves the following program.

$$\begin{aligned}
\max_{s,f} \quad & \int (B(\mathbf{x}) - s(\mathbf{x}))f(\mathbf{x})d\mathbf{x} \\
\text{s.t.} \quad & \nu + \int (U(s(\mathbf{x})) - \nu)f(\mathbf{x})d\,\mathbf{x} - V(f) \geq \bar{U} \\
& U(s(\mathbf{x})) = \ln\left(\frac{f(\mathbf{x})}{g(\mathbf{x})}\right) + 1 + \nu \text{ for all } \mathbf{x} \\
& 1 = \int f(\mathbf{x})d\mathbf{x}
\end{aligned} \tag{12}$$

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 11); the first-order approach is valid here because the agent's program (10) 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 (11) must be satisfied for all $\mathbf{x}$ in order 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."[14] 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.

Pointwise optimization of (12) with respect to s at $\mathbf{x}$ yields

$$\frac{1}{U'(s(\mathbf{x}))} = \lambda + \mu(\mathbf{x})\frac{1}{f(\mathbf{x})}. \tag{13}$$

The term $1/f(\mathbf{x})$ in equation (13) is the analog to the classic likelihood ratio, $f_a(\mathbf{x})/f(\mathbf{x})$, in equation (5). To see this, notice that because the agent chooses each point in the distribution f independently, the derivative of $f(\mathbf{x})$ with respect to the agent's choice of f at the point $\mathbf{x}$ is equal to one, and hence we can compute the likelihood ratio as $\frac{\frac{d}{df}f(\mathbf{x})}{f(\mathbf{x})} = \frac{1}{f(\mathbf{x})}$. There is an important difference between equations (5) and (13), stemming from the agent's rich action space: in equation (5), there is a *single* IC multiplier, μ, while in equation (13) there is a $\mu(\mathbf{x})$ at each point $\mathbf{x}$. The following proposition shows that when we solve for $\mu(\mathbf{x})$, the term $\mu(\mathbf{x})\frac{1}{f(\mathbf{x})}$ in equation (13) is replaced by the principal's net objective, $B(\mathbf{x}) - s(\mathbf{x})$.

[14]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. To see the equivalence, 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 $1 = \int f(\mathbf{x})d\mathbf{x}$.

Hence, in stark contrast to the classic parametric approach, likelihood ratios play no role in the solution.

Proposition 1 *The contract solving program 12 is characterized as follows.*

$$\frac{1}{U'(s(\mathbf{x}))} \quad = \quad \lambda - \eta + B(\mathbf{x}) - s(\mathbf{x}). \tag{14}$$

The vector $\mathbf{x}$ enters (14) only through the contract, $s(\mathbf{x})$, and the objective, $B(\mathbf{x})$. All variation in the contract comes through variation in the objective $B(\mathbf{x})$. This has important implications for optimal aggregation. Rearranging (14) gives

$$s(\mathbf{x}) \quad = \quad \tilde{U}^{-1}\left(B(\mathbf{x})\right), \tag{15}$$

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 the elements of $\mathbf{x}$ are first aggregated into a composite measure and then a compensation contract is written on that composite measure.

$$\begin{aligned} \textit{Aggregator:} \qquad \pi(\mathbf{x}) \quad &= \quad B(\mathbf{x}) \\ \textit{Contract:} \qquad s(\pi(\mathbf{x})) \quad &= \quad \tilde{U}^{-1}\left(\pi(\mathbf{x})\right) \end{aligned} \tag{16}$$

Thus, the principal's objective, $B(\mathbf{x})$, is an optimal aggregator for contracting.

The reason for this result stems from the agent's rich action space. Notice that in the classic approach (section 3.1), the agent chooses a scalar, $a \in \mathbb{R}$, whereas the principal chooses a vector-valued function, $s(\mathbf{x})$. Because the agent's options are so limited relative to the principal's, there are infinitely many contracts that can implement a particular action a, and the principal has the advantage of choosing among these to maximize risk-sharing efficiency. The generalized approach puts the principal and agent on equal footing: the agent chooses $f(\mathbf{x})$ at every $\mathbf{x}$ and the principal chooses $s(\mathbf{x})$ at every $\mathbf{x}$.

With this balance of control, any change in the contract s causes the agent to change his choice of f. Hence, the optimal contract (14) is unique up to addition by a constant: there is only *one* contract that will implement a particular action while rewarding the agent his reservation utility. The principal wants the largest possible mean-shift in $B(\mathbf{x})$, and the one-to-one mapping between the contract and the distribution implies that any attempt by the principal to weight the elements of $\mathbf{x}$ by some scheme $\delta(\mathbf{x})$ rather than $B(\mathbf{x})$ would result in the agent changing his choice of $f(\mathbf{x})$ to improve $\delta(\mathbf{x})$ rather than $B(\mathbf{x})$ (akin to the agent "misallocating effort" among the components that the principal cares about).

Banker and Datar [1989] provided a rationale for linear aggregation by showing that under the classic approach, optimal aggregators can be linear if likelihood ratios are linear. By contrast, the optimal aggregator in our setting is linear if the principal's objective $B(\mathbf{x})$ is linear. The optimal weights in the linear aggregator are determined entirely by the linear weights in $B(\mathbf{x})$, 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.*

$$\begin{aligned} \text{Aggregator:} \qquad \pi(\mathbf{x}) &\equiv \sum_{i=1}^{m} \pi_i x_i = \sum_{i=1}^{m} b_i x_i \\ \text{Contract:} \qquad s(\pi(\mathbf{x})) &= \tilde{U}^{-1}\left(\pi(\mathbf{x})\right), \end{aligned} \tag{17}$$

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 $\mathbf{b}$.

Corollary 1 can help rationalize the use of equally-weighted aggregates in compensation contracts. As an example, assume that $\mathbf{x} = (x_1, x_2)^T$, where x_1 is income from Product 1 and x_2 is income from Product 2. Assume that the principal values income from these products equally such that her objective can be written $B(x_1, x_2) = x_1 + x_2$. Assume further that income from one of the products is noisier that the other because, for example, its demand is more elastic. Then the classic sensitivity-precision result from Banker and Datar [1989] would predict a lower contractual weight on the noisier product (all else equal), and that it would therefore be inefficient to contract on total income, $x_t \equiv x_1 + x_2$. Under our approach, by contrast, the optimal contractual aggregator follows from (17) as

$$\pi(x_1, x_2) = B(x_1, x_2) = x_1 + x_2 = x_t. \tag{18}$$

That is, contracting on total income is perfectly efficient, despite the differing degrees of noisiness of the underlying products.

4 Full model: Contracting on aggregated estimates

In the prior section, we assumed that $\mathbf{x}$ was observable; this allowed us to make straightforward comparisons to Banker and Datar [1989], who also assume that $\mathbf{x}$ is observable. But if the things firm owners value were observable, there would be no need for accounting measurement. In the language suggested by Ijiri [1967], accounting numbers are *surrogates* (things that represent other things or phenomena) that represent certain economic *principals* (things or phenomena represented by surrogates). We now extend the model to a setting where the things firm owners care about are unobservable, which will allow us to jointly study measurement and aggregation.

Assume as before that the principal cares about m constructs, $\mathbf{x} \equiv (x_1, \ldots, x_m)^T$, and assume that she values these constructs linearly such that her objective is $B(\mathbf{x}) = \sum_{i=1}^{m} b_i x_i$. Now assume that the principal cannot contract on $\mathbf{x}$; this is in line with the idea from accounting thought that notions such as "income" are abstract constructs that cannot be directly observed (Ijiri [1975]). Instead, the principal can observe n contractible random variables given by $\mathbf{y} \equiv (y_1, \ldots, y_n)^T$. Importantly, $\mathbf{y}$ must be verifiable in order to form the basis of a legally enforceable contract. We interpret $\mathbf{y}$ as the entire set of verifiable events and transaction characteristics relevant to the contracting period. This includes all exchanges with other parties—the inflowing and outflowing "facts of services rendered" described by Paton and Littleton [1940]—and accompanying data such as inventory delivery times or customer credit scores.

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 define the cost of implementing $f(\mathbf{x}, \mathbf{y})$ by its divergence from an exogenous reference distribution, $g(\mathbf{x}, \mathbf{y}) \in \Delta(\mathbf{x}, \mathbf{y})$:

$$V(f) \quad = \quad D(f(\mathbf{x}, \mathbf{y})||g(\mathbf{x}, \mathbf{y})) \quad \equiv \quad \int f(\mathbf{x}, \mathbf{y}) \ln \left(\frac{f(\mathbf{x}, \mathbf{y})}{g(\mathbf{x}, \mathbf{y})} \right) d(\mathbf{x}, \mathbf{y}), \tag{19}$$

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. The cost-minimizing distribution $g(\mathbf{x}, \mathbf{y})$ captures "natural relationships" between fundamentals $\mathbf{x}$ and observables $\mathbf{y}$. For example, a natural relationship between the number of sales transactions (observable, in $\mathbf{y}$) and the increase in net economic assets (unobservable, in $\mathbf{x}$) would be captured by these variables being correlated in $g(\mathbf{x}, \mathbf{y})$, indicating that it is costly for the agent to increase sales without also increasing net economic assets.

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.

$$\begin{array}{ll} \max\limits_{f} & \int U(s(\mathbf{y}))f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) - \int \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} & 1 = \int f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \end{array} \tag{20}$$

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

$$f(\mathbf{x},\mathbf{y}) \;=\; g(\mathbf{x},\mathbf{y})e^{U(s(\mathbf{y}))-1-\nu} \tag{21}$$

Because $\mathbf{x}$ is not contractible, the agent has 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.

$$\begin{array}{ll} \max\limits_{s,f} & \int (B(\mathbf{x}) - s(\mathbf{y}))f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} & \nu + \int \left(U(s(\mathbf{y})) - \nu\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) - \int \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \geq \bar{U} \\ & U(s(\mathbf{y})) = \ln\left(\frac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}\right) + 1 + \nu \text{ for all } (\mathbf{x},\mathbf{y}) \\ & 1 = \int f(\mathbf{x},\mathbf{y})d(\mathbf{x},\mathbf{y}) \end{array} \tag{22}$$

Due to the principal's control disadvantage, there are some distributions that she cannot implement with any contract. Pointwise optimization of program (22) 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 (22).

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.*

$$\begin{array}{rcl} \frac{1}{U'(s(\mathbf{y}))} & = & \lambda - \eta + \int B(\mathbf{x})f(\mathbf{x}\,|\,\mathbf{y})d\,\mathbf{x} - s(\mathbf{y}) \\ \iff s(\mathbf{y}) & = & \tilde{U}^{-1}\left(\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}]\right), \end{array} \tag{23}$$

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 $\tilde{U}(s) \equiv s + \frac{1}{U'(s)} - \lambda + \eta$ characterizes the contract's functional form.

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 $\mathbf{y}$. Then contracting on $\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}]$ is just as efficient as contracting on all of the data $\mathbf{y}$; in other words, $\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}]$ is an efficient aggregator of the information $\mathbf{y}$.

Closer examination reveals that this aggregator can be separated very naturally into two functional components. 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:

$$\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}] \;=\; \mathbb{E}_f\left[\sum_{i=1}^m b_i x_i \middle|\, \mathbf{y}\right] \;=\; \sum_{i=1}^m b_i\, \mathbb{E}_f[x_i|\,\mathbf{y}], \tag{24}$$

where $\mathbb{E}_f[x_i|\,\mathbf{y}] \equiv \int x_i f(x_i|\,\mathbf{y}) dx_i$ is the expected value of x_i given all available information in $\mathbf{y}$; that is, it is an *estimate* of x_i given the data $\mathbf{y}$. Equation (24) reveals that one very natural way to construct the optimal aggregate $\mathbb{E}_f[B(\mathbf{x})|\,\mathbf{y}]$ is to first estimate each element in $\mathbf{x}$ and then aggregate the estimates by the linear weights $\mathbf{b}$. This idea is formalized in the following corollary, where we separate the principal's solution into three functional components: estimation, aggregation, and compensation. We let $\hat{x}_i$ denote the principal's estimate of x_i and $\hat{\mathbf{x}}$ 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 estimate, aggregator, and compensation contract are as follows.*

$$\begin{array}{rrcl} \text{Estimates:} & \hat{\mathbf{x}} & = & \mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}] \\ \text{Aggregator:} & \pi(\hat{\mathbf{x}}) & = & B(\hat{\mathbf{x}}) \equiv \mathbf{b}^T \hat{\mathbf{x}} \\ \text{Contract:} & s(\pi(\hat{\mathbf{x}})) & = & \tilde{U}^{-1}\left(\pi(\hat{\mathbf{x}})\right), \end{array} \tag{25}$$

where $\mathbb{E}_f[\mathbf{x}\,|\,\mathbf{y}] \equiv \int \mathbf{x}\, f(\mathbf{x}\,|\,\mathbf{y}) d\,\mathbf{x}$ is the expected value of $\mathbf{x}$ given $\mathbf{y}$ under the equilibrium distribution f 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. In the *measurement* stage, the principal estimates each x_j as $\hat{x}_j = \mathbb{E}_f[x_j|\,\mathbf{y}]$. Notice that $\hat{x}_j$ is an *unbiased estimate*:

$$\mathbb{E}_f[\hat{x}_j] = \mathbb{E}_f[\mathbb{E}_f[x_j|\,\mathbf{y}]] = \mathbb{E}_f[x_j], \tag{26}$$

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]). In the *aggregation* stage, the principal linearly aggregates the estimates $\hat{\mathbf{x}}$ according to the weights in her objective to produce the aggregated estimate $\pi(\hat{\mathbf{x}}) = \mathbf{b}^T\hat{\mathbf{x}} = \sum_{i=1}^{m} b_i\hat{x}_i$. Finally, in the *compensation* stage, the principal conditions the agent's compensation on the aggregated estimate.

Our decomposition mirrors how accounting is done in practice. Standard setters define unobservable constructs such as revenues and expenses and dictate rules for how those constructs are to be measured. This process is generally highly non-linear and contains many contingencies, such as criteria for revenue recognition. The resulting metrics are then aggregated in a process that is strictly linear: under double-entry conventions, individual transactions and accounts are aggregated with weights of 1 or -1 to form aggregated metrics. Executives are then compensated on one or more of these aggregated metrics. Our results indicate that contracting on a given single accounting aggregate in practice is efficient if 1) the component accounting proxies are representationally faithful (i.e. $\hat{\mathbf{x}}$ is an unbiased estimate of $\mathbf{x}$) and 2) the principal values the underlying constructs $\mathbf{x}$ with weights of 1 or -1, just as those elements are aggregated in conventional accounting.

To illustrate, assume that the principal cares equally about the unobservable constructs of FASB-defined revenues, x_r, and FASB-defined expenses, x_e, such that her objective is $B(x_r, x_e) = x_r - x_e \equiv z$, where z denotes true (unobservable) earnings. The principal cannot observe x_r, x_e or z directly, but she 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 write a highly complex contract conditioned on all of the underlying data in $\mathbf{y}$ (indeed, compensation committees have access to enormous databases of detailed information). 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 $\hat{x}_e$:

$$\begin{array}{lrcl} \textit{Estimate of } x_r & \hat{x}_r & = & \mathbb{E}_f[x_r|\,\mathbf{y}] \\ \textit{Estimate of } x_e & \hat{x}_e & = & \mathbb{E}_f[x_e|\,\mathbf{y}]. \end{array} \tag{27}$$

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

$$\begin{array}{rrcl} \textit{Aggregator:} & \pi(\hat{x}_r, \hat{x}_e) & = & B(\hat{x}_r, \hat{x}_e) = \hat{x}_r - \hat{x}_e \equiv \hat{z} \\ \textit{Contract:} & s(\pi(\hat{x}_r, \hat{x}_e)) & = & s(\hat{z}) = \tilde{U}^{-1}\left(\hat{z}\right). \end{array} \tag{28}$$

This example illustrates that contracting solely on net income in practice is efficient to the extent that 1) the constructs of revenues and expenses are defined by changes in shareholder wealth and 2) measured revenues and expenses ($\hat{x}_r$ and $\hat{x}_e$) faithfully represent the FASB-defined constructs they purport to represent (x_r and x_e). Under these conditions, contracting solely on measured net income is 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 multiple accounting metrics. One commonly observed practice in executive compensation is to condition CEO pay on both revenue and income metrics (De Angelis and Grinstein [2015], Bloomfield, Gipper, Kepler, and Tsui [2021]). Extending the example above, suppose that the principal puts a higher valuation weight on revenues than expenses because, for example, the FASB-defined construct of revenues does not fully capture changes in shareholder wealth. Specifically, assume that the principal's objective is given by $B(x_r, x_e) = b_r x_r - x_e$, with $b_r > 1$. In this case, contracting solely on measured net income ($\hat{z} \equiv \hat{x}_r - \hat{x}_e$) is not efficient. Instead, the optimal contract can be conditioned on a weighted aggregate of estimated revenues and expenses,

$$\pi(\hat{x}_r, \hat{x}_e) = b_r \hat{x}_r - \hat{x}_e, \tag{29}$$

or equivalently, on revenues and net income:

$$\pi(\hat{x}_r, \hat{z}) = (b_r - 1)\hat{x}_r + \hat{z}, \tag{30}$$

where $b_r - 1 > 0$. Contracts that put positive incentives on revenues and earnings are in effect weighting revenues more heavily than expenses, and our results suggest this is done to align the contractual weights with the principal's objective.

5 Optimal measurement rules

Corollary 2 shows that the optimal contract is conditioned on the unbiased estimates $\hat{\mathbf{x}} = \mathbb{E}_f[\mathbf{x}|\mathbf{y}]$. As in Gao [2013], there is a two-step mapping from the fundamentals $\mathbf{x}$ to the performance measures $\hat{\mathbf{x}}$. First, as part of the agent's choice of $f(\mathbf{x}, \mathbf{y})$, the unobservable fundamentals $\mathbf{x}$ map to the transaction characteristics $\mathbf{y}$ according to the conditional distribution $f(\mathbf{y}|\mathbf{x})$. Then, as part of the principal's solution, the transaction characteristics $\mathbf{y}$ are used to form the estimates $\hat{\mathbf{x}}$. We will use the term *measurement* to refer to the second part of this two step mapping, where the principal writes measurement rules on observable transaction characteristics. We will refer to *optimal measurement rules* as functions of $\mathbf{y}$ that produce unbiased estimates of $\mathbf{x}$. Notice that under the general model in the prior section, the measurement process is ambiguous; without knowing the form of the distribution f, we cannot say very much about how the principal forms the expectation $\mathbb{E}_f[\mathbf{x}|\mathbf{y}]$. In this section, we employ a 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 (19). 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 $U(s) = s$. The assumption of risk neutrality does not make the agency problem trivial; when the principal cannot contract on the outcomes she values (here, $\mathbf{x}$), she cannot achieve first-best by selling the firm to a risk-neutral agent (Baker [1992]). We are giving up risk aversion for tractability, 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 his preferred distribution $g(\mathbf{x}, \mathbf{y})$ 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.*

$$\begin{array}{rrcl} \text{Estimate:} & \hat{\mathbf{x}} & = & \Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y} \\ \text{Aggregator:} & \pi(\hat{\mathbf{x}}) & = & B(\hat{\mathbf{x}}) = \mathbf{b}^T\hat{\mathbf{x}} \\ \text{Contract:} & s(\pi(\hat{\mathbf{x}})) & = & \pi(\hat{\mathbf{x}}) - \eta \\ \text{Action:} & f(\mathbf{x}, \mathbf{y}) & = & g(\mathbf{x}, \mathbf{y})e^{\mathbf{b}^T\Sigma_{\mathbf{xy}}\Sigma_{\mathbf{yy}}^{-1}\mathbf{y} - \eta}, \end{array} \tag{31}$$

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 $\mu_{\mathbf{y}} = \Sigma_{\mathbf{yx}}\mathbf{b}$. Moreover, the equilibrium cost $V(f)$ is increasing quadratically in the mean shift from g to f.

The solution (31) is presented using a three-stage decomposition, as in Corollary 2, where the principal first constructs the estimates $\hat{x}_1 \dots \hat{x}_m$, 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 $s(\mathbf{y}) = \mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y}$.

The matrices $\Sigma_{\mathbf{x}\,\mathbf{y}}$ and $\Sigma_{\mathbf{y}\,\mathbf{y}}^{-1}$ in the unbiased estimator are analogous to the classic notions of sensitivity and precision, where we can interpret $\Sigma_{\mathbf{y}\,\mathbf{y}}^{-1}$ as the precision of the measures $\mathbf{y}$ and $\Sigma_{\mathbf{x}\,\mathbf{y}}$ as the sensitivity of the measures $\mathbf{y}$ to changes in the constructs $\mathbf{x}$. Hence, the term $\Sigma_{\mathbf{x}\,\mathbf{y}} \Sigma_{\mathbf{y}\,\mathbf{y}}^{-1}$ can be interpreted as the "signal-to-noise" ratios of the measures $\mathbf{y}$ with respect to the *constructs* $\mathbf{x}$. This is distinct from Banker and Datar [1989], where the signal-to-noise ratio $\frac{\partial E(x_i|a)/\partial a}{\mathrm{Var}(x_i)}$ is defined for the variable x_i with respect to the *agent's action*. Notice from equation (31) that sensitivity $\Sigma_{\mathbf{x}\,\mathbf{y}}$ and precision $\Sigma_{\mathbf{y}\,\mathbf{y}}^{-1}$ enter the solution only through forming the estimates $\hat{\mathbf{x}}$. Crucially, when $\mathbf{x}$ is observable (i.e. $\hat{\mathbf{x}} = \mathbf{x}$), sensitivity and precision play *no role* in the solution (see also Corollary 1). This stands in stark contrast to Banker and Datar [1989], where $\mathbf{x}$ is assumed to be observable and the principal aggregates the elements of $\mathbf{x}$ according to their precision and sensitivity to the agent's action (as we reviewed in section 3.1).

Equation (31) shows that our specification in this section changes the general solution (25) in three ways. First, measurement is no longer ambiguous; the principal's unbiased estimate is given by the linear regression $\hat{\mathbf{x}} = E_f[\mathbf{x}|\mathbf{y}] = \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y}$.[15] Second, the optimal contract is now linear in the estimates $\hat{\mathbf{x}}$. This linearity comes from assuming risk neutrality. While risk neutrality produces weakly optimal linear contracts under the classic approach, here it results in *uniquely* optimal linear contracts. Finally, the tractable specification produces 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 from g but the same variance-covariance matrix.

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

[15] 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 $f(\mathbf{x}, \mathbf{y}) = f(\mathbf{x}|\mathbf{y}) f(\mathbf{y})$.

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.

Interestingly, our solution has several properties that are similar to the assumptions made in the LEN model (e.g. Feltham and Xie [1994]). LEN models exogenously restrict contracts to be linear; our solution has uniquely optimal linear contracts. In LEN, the agent is modeled as choosing the mean of a normal distribution; for example, the agent chooses a in $f \sim \mathcal{N}(a, \sigma^2)$. In our solution, the agent shifts the mean of a normal distribution. LEN models often assume that the agent's personal cost is quadratic in the mean shift a; our Proposition 3 shows that the agent's equilibrium cost is quadratic in the implemented mean shift. The LEN assumptions are made for tractability, and our specification produces many of the same tractable features.

One notable difference between our specification and LEN is that we assume risk neutrality while the LEN agent has negative exponential utility. Therefore, the specification we use in this section is not equipped for studying risk-sharing issues. The LEN model sacrifices optimal contracts to gain tractability in complex settings, and we give up risk considerations to study optimal measurement as well as the effects of measurement on production. However, optimal measurement in our setting does not depend on the agent being risk neutral; we show in Appendix B that our main measurement insights are robust risk aversion.

What the risk neutrality assumption buys us, in conjunction with the assumption that $g(\mathbf{x}, \mathbf{y})$ is a multivariate normal, is that $f(\mathbf{x}, \mathbf{y})$ is multivariate normal in equilibrium. The normality of $f(\mathbf{x}, \mathbf{y})$ conveniently allows for comparative statics on how the covariance structure Σ_g affects production. As a simple example, consider the case with $m = n = 1$; that is, the principal cares about a single construct (e.g. income) 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 as follows.

$$\begin{aligned} \textit{Estimate:} \quad & \hat{x} &=& \ \rho y \\ \textit{Contract:} \quad & s(\hat{x}) &=& \ \hat{x} - \tfrac{1}{2}\rho^2 \\ \textit{Action:} \quad & f(x, y) &\sim& \ \mathcal{N}\left(\left[\begin{smallmatrix} \rho^2 \\ \rho \end{smallmatrix}\right], \Sigma_g\right) \end{aligned} \tag{32}$$

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 might also interpret ρ as an index of the *congruity* between x and y, where x and y are perfectly congruous when $\rho = 1$ and totally incongruous when $\rho = 0$. As in Feltham and Xie [1994],[16] first best allocations are attainable with a risk neutral agent only when y is perfectly congruent.

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. Although these applications are highly stylized, they do suggest that many common accounting practices are efficient for contracting when managers have extensive control over both fundamental performance and observable data.

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 improve 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 improving the observable transaction characteristic y but not the unobservable true performance x. 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.

[16] In Feltham and Xie [1994], the principal's objective x and the performance measure y are both linear in a vector of the agent's efforts along different dimensions, plus some normally-distributed noise: $x = \mathbf{b}^T\mathbf{a} + \epsilon_x$ and $y = \mathbf{q}^T\mathbf{a} + \epsilon_y$. In their setting, *congruity* refers to proportionality between the vectors $\mathbf{b}$ and $\mathbf{q}$; a perfectly congruous performance measure y weights the dimensions of effort the same way that the principal's objective does.

Assume that the agent's preferred distribution, $g(x, y)$, is a bivariate normal with mean $\mathbf{0}$ and variance-covariance matrix $\Sigma_g = \left[\begin{smallmatrix} 1 & \rho \\ \rho & 1 \end{smallmatrix}\right]$, where $\rho > 0$. 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 y 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 about true performance. 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 to the principal's program is given by equation (32). Examining the equilibrium distribution reveals 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 (32), 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], the conservative measurement rule offsets managerial biases to create an unbiased performance measure.

5.2 Accounting for investments

Let the principal's objective be Hicksian income (the change in net economic assets during the period), defined as $B(\mathbf{x}) = b_2 x_2 - x_1$, where x_1 represents current outlays in an investment and $b_2 x_2$ represents future benefits from the investment, where $b_2 > 1$ so that the investment is expected to have positive NPV. 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

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_{x_1x_2} & \rho_{x_1y_1} & \rho_{x_1y_2} \\ \rho_{x_1x_2} & 1 & \rho_{x_2y_1} & \rho_{x_2y_2} \\ \rho_{x_1y_1} & \rho_{x_2y_1} & 1 & \rho_{y_1y_2} \\ \rho_{x_1y_2} & \rho_{x_2y_2} & \rho_{y_1y_2} & 1 \end{bmatrix}. \tag{33}$$

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 (33) can be rewritten as

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho_x & 1 & \rho_y \\ \rho_x & 1 & \rho_x & \rho_2 \\ 1 & \rho_x & 1 & \rho_y \\ \rho_y & \rho_2 & \rho_y & 1 \end{bmatrix}. \tag{34}$$

From Proposition 3, the solution is as follows.

$$\begin{aligned} &\textit{Estimate of } x_1: && \hat{x}_1 = y_1 \\ &\textit{Estimate of } x_2: && \hat{x}_2 = \left(\tfrac{\rho_x - \rho_y\rho_2}{1-\rho_y^2}\right) y_1 + \left(\tfrac{\rho_2 - \rho_x\rho_y}{1-\rho_y^2}\right) y_2 \\ &\textit{Aggregator}: && \pi(\hat{\mathbf{x}}) = B(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 \\ &\textit{Contract}: && s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\ &\textit{Action}: && f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = \rho_x b_2 - 1 \\ \mu_{x_2} = \frac{b_2(\rho_x^2 + \rho_2^2 - 2\rho_x\rho_y\rho_2) - \rho_x(1-\rho_y^2)}{1-\rho_y^2} \\ \mu_{y_1} = \rho_x b_2 - 1 \\ \mu_{y_2} = \rho_2 b_2 - \rho_y \end{bmatrix}, \Sigma_g\right) \end{aligned} \tag{35}$$

To gain intuition for this solution, we will do comparative statics under two cases, one in which reliable evidence about 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: $\rho_{x_1y_2} = \rho_{x_2y_2} = \rho_{y_1y_2} = 0$. With this assumption, the variance-covariance matrix under the agent's cost-minimizing distribution g is given as follows, where

$\rho_x \equiv \rho_{x_1x_2} = \rho_{x_2y_1}$.

$$\Sigma_g \quad = \quad \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \quad = \quad \begin{bmatrix} 1 & \rho_x & 1 & 0 \\ \rho_x & 1 & \rho_x & 0 \\ 1 & \rho_x & 1 & 0 \\ 0 & 0 & 0 & 1 \end{bmatrix}. \tag{36}$$

Under these assumptions, the solution (35) reduces to the following.

$$\begin{aligned} \textit{Estimate of } x_1\text{:} \quad & \hat{x}_1 = y_1 \\ \textit{Estimate of } x_2\text{:} \quad & \hat{x}_2 = \rho_x y_1 \;=\; \rho_x \hat{x}_1 \\ \textit{Aggregator:} \quad & \pi(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 \;=\; (b_2\rho_x - 1)\hat{x}_1 \\ \textit{Contract:} \quad & s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\ \textit{Action:} \quad & f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = b_2\rho_x - 1 \\ \mu_{x_2} = \rho_x(b_2\rho_x - 1) \\ \mu_{y_1} = b_2\rho_x - 1 \\ \mu_{y_2} = 0 \end{bmatrix}, \Sigma_g \right) \end{aligned} \tag{37}$$

Both $\hat{x}_1$ and $\hat{x}_2$ are unbiased measures of x_1 and x_2 given the equilibrium action: $\mathbb{E}_f[\hat{x_1}] = \mathbb{E}_f[y_1] = b_2\rho_x - 1 = \mu_{x_1} = \mathbb{E}_f[x_1]$ and $\mathbb{E}_f[\hat{x_2}] = \rho_x \mathbb{E}_f[y_1] = \mu_{x_2} = \mathbb{E}_f[x_2]$. Interpret the aggregation function $\pi(\hat{\mathbf{x}}) = (b_2\rho_x - 1)\hat{x}_1$ as reported net income pertaining to the investment, and note that this is increasing in ρ_x for all positive investments. Thus, the optimal measurement rule depends on the natural correlation between current investment and future returns.

If $\rho_x = 0$ so that investments are maximally uninformative about future returns, then net income is given by the cash outlay: $\pi(\hat{\mathbf{x}}) = -\hat{x}_1 = -y_1$. That is, investment spending is immediately expensed and no future benefits are estimated. This is loosely analogous to the treatment of R&D or advertising expenditures, which generally have a fuzzy mapping to future returns and are immediately expensed under U.S. GAAP. 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. Notice that if net income were to fully accrue the unrealized income and naively ignore $\rho_x = 0$ (i.e., if $\pi = (b_2 - 1)y_1$), the manager would respond by making excessive R&D investments that are unlikely to produce future returns. Optimal measurement of x_2 ($\hat{x}_2 = \rho_x y_1$) discounts y_1 by ρ_x so that the manager does not game investments to inflate reported income. Hence, although conservative reporting of future investment income causes the manager to cut R&D, it mitigates overinvestment in bad projects whose future returns will never materialize in expectation.

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 not reflected on the income statement; in a double-entry system, the investment is 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 $\pi(\hat{\mathbf{x}}) = (b_2 - 1)\hat{x}_1$. A useful analogy is the sacrifice of inventory in a credit sale, which tends to be highly correlated with the future receipt of cash from customers. In that case, 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. With $\rho_x = 1$, it is very difficult for the manager to make inventory deliveries that are not accompanied by future cash collections; said differently, delivering inventory and collecting the corresponding payments are natural complements in the agent's cost function. 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. In this limiting case where $\rho_x = 1$, conservatism is unnecessary in the optimal measurement rule because it is very difficult for the manager to engage in real activities manipulation. With ρ_x slightly less than one, we could think of the manager as being able to game investments on the margins, such as engaging in channel stuffing behavior or delivering inventory to customers with bad credit. In these situations, the optimal measurement rule would not fully accrue the future income; for example, it might offset some accrued revenue with bad debt expense.

These rules show that when future returns are difficult to measure (which we contend is the typical case), optimal measurement of investments depends on the likelihood that the investment will pay off. This is consistent with the idea that accruals contain implicit forecasts or assumptions about the future (Beaver [1991], Leuz [1998], Ordelheide [1988]). Next we examine optimal measurement when future returns are easy to measure.

Case 2: Future returns easy to measure

Now we relax the assumption that y_2 is pure noise and consider how the solution (35) 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.[17] Then setting $\rho_2 = 1$ and letting $\rho \equiv \rho_x = \rho_y$, the variance-covariance matrix is as follows.

$$\Sigma_g = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} = \begin{bmatrix} 1 & \rho & 1 & \rho \\ \rho & 1 & \rho & 1 \\ 1 & \rho & 1 & \rho \\ \rho & 1 & \rho & 1 \end{bmatrix}. \tag{38}$$

This reduces solution (35) to the following.

$$\begin{aligned} \textit{Estimate of } x_1\text{:} \quad & \hat{x}_1 = y_1 \\ \textit{Estimate of } x_2\text{:} \quad & \hat{x}_2 = y_2 \\ \textit{Aggregator:} \quad & \pi(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 \\ \textit{Contract:} \quad & s(\pi(\hat{\mathbf{x}})) = \pi(\hat{\mathbf{x}}) - \eta \\ \textit{Action:} \quad & f(x_1, x_2, y_1, y_2) \sim \mathcal{N}\left(\begin{bmatrix} \mu_{x_1} = \rho b_2 - 1 \\ \mu_{x_2} = b_2 - \rho \\ \mu_{y_1} = \rho b_2 - 1 \\ \mu_{y_2} = b_2 - \rho \end{bmatrix}, \Sigma_g \right) \end{aligned} \tag{39}$$

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 Case 1, conservatism is not needed to construct an unbiased estimate of future returns. Recall that the principal's objective is $B(\mathbf{x}) = b_2x_2 - x_1$, and suppose that $b_2 > 1$ represents the principal's discount factor on the perpetuity 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 firm wealth from the investment is reported at *fair value*, the present value of expected future cash flows less the amount expended in the current period.

The results from Case 1 and 2 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 is driven by ρ_x, the correlation between current investments

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

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 is available, optimal measurement rules take that evidence into account (see $\hat{x}_2$ in equation 35), 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

We provide a framework for studying the contracting role of accounting. In this framework, a manager has nonparametric influence over the joint distribution of unobservable fundamentals ($\mathbf{x}$) and observable events and transaction characteristics ($\mathbf{y}$). Accounting is the set of *measurement* and *aggregation* rules that estimates $\mathbf{x}$ using observable data $\mathbf{y}$ and then aggregates those estimates to produce one or more aggregated estimates.

Some elements of $\mathbf{x}$ may be unobservable because they are future events not realized in the contracting period. Accounting measurement is thus both *forward looking* and *backward looking*: it forecasts future outcomes ($\mathbf{x}$) using past transactions and events ($\mathbf{y}$). This pulls together ideas from accounting thought that accounting measurement should rely on verifiable "backward-looking" information to facilitate legally enforceable contracts and to hold managers accountable (e.g. Butterworth, Gibbins, and King [1982], Ijiri [1975]) and that accounting measures contain implicit assumptions about the future (e.g. Ordelheide [1988], Beaver [1991], Leuz [1998]).

As such, our measurement framework rectifies a long-standing debate in accounting thought, described by George O. May in 1943:

> [T]he present ferment in accounting thought is very largely due to the conflicting objectives of those who would continue to regard financial statements as reports of progress or of stewardship, and those who would treat them as being in the nature of prospectus. (May [1943], p. 21.)

The objectives described by May do not conflict in our model. The optimal measures $\hat{\mathbf{x}}$ are both contractually optimal, satisfying the stewardship objective (e.g. Watts and Zimmerman [1986], Ijiri [1975]), and are faithful representations (i.e. unbiased estimates) of the underlying economic phenomena $\mathbf{x}$, satisfying what Zeff [2013] termed the "representationalist" objective (e.g. Moonitz [1961]).

In our framework, we separate the principal's solution into three stages: 1) using verifiable data $\mathbf{y}$ to estimate $\mathbf{x}$, 2) aggregating the resulting estimates $\hat{\mathbf{x}}$, and 3) compensating the agent based on the aggregated estimate $\pi(\mathbf{x})$. This process is descriptive of practice and provides a rationale for the widespread use of aggregated accounting metrics in executive compensation contracts. Compensation committees have access to enormous databases of information (e.g. individual exchange transactions, inventory delivery times, and customer credit scores), but rather than referencing all of this information, executive compensation contracts are conditioned on a handful of aggregated accounting metrics. Our results indicate that this practice is optimal to the extent that 1) GAAP constructs are defined as changes in shareholder wealth (such that shareholders value the GAAP constructs equally) and 2) GAAP measurement rules produce metrics that faithfully represent the underlying GAAP constructs.

In section 5, we provide a tractable specification of our framework for studying optimal accounting measurement. Here too, the results are very descriptive. We find that optimal measurement is conservative: because managers can game evidence $\mathbf{y}$ about unobservable fundamentals $\mathbf{x}$, the evidence is discounted so that estimated fundamentals $\hat{\mathbf{x}}$ are unbiased. We also provide a measurement rule for accounting for uncertain investments: the degree to which unrealized income is recognized should depend on the natural relationship between current investment and future returns.

The tractable specification that we develop in section 5 has *in equilibrium* many of the same features that the LEN model assumes *ex ante*. Our specification also has the advantage of producing optimal contracts (rather than linearly restricted ones). Much like how the LEN model opened the door to studying applied risk-sharing issues in a linear contracting setting, we hope our framework will facilitate the study of measurement issues in an optimal contracting setting.

A Proofs

Proof of Proposition 1. First, write program (12) in Lagrangian form as follows.

$$\begin{aligned}\mathcal{L} &= \int (B(\mathbf{x}) - s(\mathbf{x})) f(\mathbf{x}) d\mathbf{x} \\ &+\lambda \left[\nu + \int (U(s(\mathbf{x})) - \nu) f(\mathbf{x}) dx - V(f) - \bar{U}\right] \\ &+ \int \mu(\tilde{\mathbf{x}}) \left[U(s(\tilde{\mathbf{x}})) - \ln\left(\tfrac{f(\tilde{\mathbf{x}})}{g(\tilde{\mathbf{x}})}\right) - 1 - \nu\right] d\tilde{\mathbf{x}} \\ &+\eta \left[1 - \int f(\mathbf{x}) d\mathbf{x}\right]\end{aligned} \tag{40}$$

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

$$\tfrac{1}{U'(s(\mathbf{x}))} \quad = \quad \lambda + \mu(\mathbf{x}) \tfrac{1}{f(\mathbf{x})}. \tag{41}$$

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

$$0 \quad = \quad B(\mathbf{x}) - s(\mathbf{x}) + \lambda \left[U(s(\mathbf{x})) - \ln\left(\tfrac{f(\mathbf{x})}{g(\mathbf{x})}\right) - 1 - \nu\right] - \mu(\mathbf{x}) \tfrac{1}{f(\mathbf{x})} - \eta \tag{42}$$

Notice that the IC constraint implies that the term in brackets is equal to zero. Then rearranging (42) gives $\mu(\mathbf{x}) = f(\mathbf{x})\,(B(\mathbf{x}) - s(\mathbf{x}) - \eta)$. Substituting this into (41) produces the solution presented in the proposition.

□

Proof of Proposition 2. Rearranging (21) 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.

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

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 $s(\mathbf{y})$:

$$\textstyle\int f(\mathbf{x},\mathbf{y}) d\,\mathbf{x} \;\; = \lambda U'(s(\mathbf{y})) \int f(\mathbf{x},\mathbf{y}) d\,\mathbf{x} + U'(s(\mathbf{y})) \int \mu(\mathbf{x},\mathbf{y}) d\,\mathbf{x} \tag{44}$$

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.

$$\frac{1}{U'(s(\mathbf{y}))} \;\; = \lambda + \left(\textstyle\int \mu(\mathbf{x},\mathbf{y}) d\,\mathbf{x}\right) \frac{1}{f(\mathbf{y})}. \tag{45}$$

Pointwise optimization of (43) with respect to f at $(\mathbf{x},\mathbf{y})$ yields the following closed form expression for $\mu(\mathbf{x},\mathbf{y})$.

$$\mu(\mathbf{x},\mathbf{y}) \;\; = \;\; f(\mathbf{x},\mathbf{y}) \left(B(\mathbf{x}) - s(\mathbf{y}) - \eta\right) \tag{46}$$

Substituting $\mu(\mathbf{x},\mathbf{y})$ into (45) gives:

$$\begin{aligned} & \frac{1}{U'(s(\mathbf{y}))} = \lambda + \left(\textstyle\int \left(B(\mathbf{x}) - s(\mathbf{y}) - \eta\right) f(\mathbf{x},\mathbf{y}) d\,\mathbf{x}\right) \frac{1}{f(\mathbf{y})} \\ \iff & \frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \textstyle\int B(\mathbf{x}) \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} d\,\mathbf{x} - s(\mathbf{y}) \\ \iff & \frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \textstyle\int B(\mathbf{x}) f(\mathbf{x}\,|\,\mathbf{y}) d\,\mathbf{x} - s(\mathbf{y}). \end{aligned} \tag{47}$$

□

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.

$$\begin{aligned} &\max_f \int s(\mathbf{y}) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) - \int f(\mathbf{x},\mathbf{y}) \ln(\tfrac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}) d(\mathbf{x},\mathbf{y}) \\ &\text{s.t.} \quad 1 = \int f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{48}$$

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

$$f(\mathbf{x},\mathbf{y}) \quad = \quad g(\mathbf{x},\mathbf{y}) e^{s(\mathbf{y})-\nu-1} \tag{49}$$

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

$$\begin{aligned} \max_{s,f,\nu} \quad & \int (B(\mathbf{x}) - s(\mathbf{y})) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \\ \text{s.t.} \quad & \nu + \int (s(\mathbf{y}) - \nu) f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) - \int f(\mathbf{x},\mathbf{y}) \ln(\tfrac{f(\mathbf{x},\mathbf{y})}{g(\mathbf{x},\mathbf{y})}) d(\mathbf{x},\mathbf{y}) \geq \bar{U} \\ & f(\mathbf{x},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) e^{s(\mathbf{y})-\nu-1} \text{ for all } (\mathbf{x},\mathbf{y}) \\ & 1 = \int f(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{50}$$

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.

$$\begin{aligned} &\max_s \int (B(\mathbf{x}) - s(\mathbf{y})) e^{s(\mathbf{y})} g(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \\ &\text{s.t.} \quad 1 = \int e^{s(\mathbf{y})} g(\mathbf{x},\mathbf{y}) d(\mathbf{x},\mathbf{y}) \end{aligned} \tag{51}$$

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

$$\begin{aligned} e^{s(\mathbf{y})} \int B(\mathbf{x}) g(\mathbf{x},\mathbf{y}) d\mathbf{x} &= (s(\mathbf{y}) + 1) e^{s(\mathbf{y})} \int g(\mathbf{x},\mathbf{y}) d\mathbf{x} + \eta e^{s(\mathbf{y})} \int g(\mathbf{x},\mathbf{y}) d\mathbf{x} \\ \iff \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] &= s(\mathbf{y}) + 1 + \eta \end{aligned} \tag{52}$$

Then we can substitute $s(\mathbf{y}) = \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] - \eta - 1$ into equation (49) to obtain the equilibrium distribution,

$$f(\mathbf{x},\mathbf{y}) \quad = \quad g(\mathbf{x},\mathbf{y}) e^{\mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] - \eta - 1}. \tag{53}$$

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.

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

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

$$\begin{aligned} \hat{\mathbf{x}} &= \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} \\ \pi(\hat{\mathbf{x}}) &= \mathbf{b}^T \hat{\mathbf{x}} \\ s(\pi(\hat{\mathbf{x}})) &= B(\hat{\mathbf{x}}) - \eta - 1 \\ f(\mathbf{x}, \mathbf{y}) &= g(\mathbf{x}, \mathbf{y}) e^{\mathbf{b}^T \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \mathbf{y} - \eta - 1} \end{aligned} \tag{55}$$

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:

$$\begin{aligned} f(\mathbf{x}, \mathbf{y}) &= (2\pi)^{-\frac{n+m}{2}} \det(\Sigma)^{-\frac{1}{2}} \exp(-\tfrac{1}{2}(\mathbf{z} - \mu)^T \Sigma^{-1} (\mathbf{z} - \mu)) \\ &= g(\mathbf{x}, \mathbf{y}) \exp(\mathbf{z}^T \Sigma^{-1} \mu - \tfrac{1}{2} \mu^T \Sigma^{-1} \mu). \end{aligned} \tag{56}$$

It follows from (55) and (56) 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.

$$\begin{aligned} \tilde{\Sigma}_{\mathbf{xx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{xy}} \mu_{\mathbf{y}} &= \mathbf{0}, \\ \tilde{\Sigma}_{\mathbf{yx}} \mu_{\mathbf{x}} + \tilde{\Sigma}_{\mathbf{yy}} \mu_{\mathbf{y}} &= \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b}, \\ \eta + 1 &= \tfrac{1}{2} \mu^T \Sigma^{-1} \mu. \end{aligned} \tag{57}$$

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:

$$\mu = \begin{bmatrix} \Sigma_{\mathbf{xx}} & \Sigma_{\mathbf{xy}} \\ \Sigma_{\mathbf{yx}} & \Sigma_{\mathbf{yy}} \end{bmatrix} \begin{bmatrix} \mathbf{0} \\ \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} \end{bmatrix} = \begin{bmatrix} \Sigma_{\mathbf{xy}} \Sigma_{\mathbf{yy}}^{-1} \Sigma_{\mathbf{yx}} \mathbf{b} \\ \Sigma_{\mathbf{yx}} \mathbf{b} \end{bmatrix} \tag{58}$$

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

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

Finally, we show that in equilibrium $V(f)$ is increasing quadratic in μ, the mean shift from g to f. For two multivariate Gaussians P_1 and P_2 in $\mathbb{R}^s$,

$$D_{KL}(P_1||P_2) = \tfrac{1}{2}\left(\ln\left(\tfrac{\det \Sigma_2}{\det \Sigma_1}\right) - s + \mathrm{tr}(\Sigma_2^{-1}\Sigma_1) + (\mu_1 - \mu_2)^T \Sigma_2^{-1}(\mu_1 - \mu_2)\right) \tag{59}$$

Then the KL divergence from $g \sim \mathcal{N}(\mathbf{0}, \Sigma_g)$ to the equilibrium distribution $f \sim \mathcal{N}(\mu, \Sigma_g)$ is computed as

$$D_{KL}(f||g) = \tfrac{1}{2}\mu^T \Sigma_g^{-1} \mu. \tag{60}$$

□

B Robustness

In this section, we show that the central insights from our measurement applications hold when the agent is risk averse. Revert to the assumptions in section 4, wherein the agent's utility function $U(s)$ is increasing concave. Our analysis builds on Proposition 2, which gives

$$\frac{1}{U'(s(\mathbf{y}))} = \lambda - \eta + \int B(\mathbf{x}) f(\mathbf{x}|\mathbf{y}) d\mathbf{x} - s(\mathbf{y}). \tag{61}$$

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

$$f(\mathbf{y}) = \int f(\mathbf{x},\mathbf{y}) dx = \int g(\mathbf{x},\mathbf{y}) e^{U(s(\mathbf{y}))-1-\nu} dx = g(\mathbf{y}) e^{U(s(\mathbf{y}))-1-\nu}, \tag{62}$$

where the second equality follows from equation (21) and the first and third equalities are by definition of a marginal distribution. Equation (62) implies that

$$f(\mathbf{x}|\mathbf{y}) = \frac{f(\mathbf{x},\mathbf{y})}{f(\mathbf{y})} = \frac{g(\mathbf{x},\mathbf{y}) e^{U(s(\mathbf{y}))-1-\nu}}{g(\mathbf{y}) e^{U(s(\mathbf{y}))-1-\nu}} = \frac{g(\mathbf{x},\mathbf{y})}{g(\mathbf{y})} = g(\mathbf{x}|\mathbf{y}). \tag{63}$$

Then $\int B(\mathbf{x}) f(\mathbf{x}|\mathbf{y}) d\mathbf{x} = \int B(\mathbf{x}) g(\mathbf{x}|\mathbf{y}) d\mathbf{x} \equiv \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}]$ and we can rewrite (61) as

$$\frac{1}{U'(s(\mathbf{y}))} + s(\mathbf{y}) = \mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] + \lambda - \eta. \tag{64}$$

Let $\alpha(s(\mathbf{y})) \equiv \frac{1}{U'(s(\mathbf{y}))} + s(\mathbf{y})$. Then $s(\mathbf{y}) = \alpha^{-1}(\mathbb{E}_g[B(\mathbf{x})|\mathbf{y}] + \lambda - \eta)$. Substituting this into the agent's first-order condition (equation 21) gives the following characterization of the incentive compatible distribution.

$$f(\mathbf{x},\mathbf{y}) = g(\mathbf{x},\mathbf{y}) e^{U(\alpha^{-1}(\mathbb{E}_g[B(\mathbf{x})|\mathbf{y}]+\lambda-\eta))-1-\nu} \tag{65}$$

B.1 Conservatism as a response to window dressing

Let the agent choose $f(x,y)$ and let the principal have objective $B(x) = x$, where y is contractible and x is not. The main insight from section 5.1 is that the agent's ability to *window dress* by improving y without improving x results in the principal designing a measurement rule that discounts the evidence y when estimating x; that is, the estimate $\mathbb{E}_f[\hat{x}]$ is less than the evidence $\mathbb{E}_f[y]$. From Corollary 2, $\hat{x} = \mathbb{E}_f[x|y]$, and thus $\mathbb{E}_f[\hat{x}] = \mathbb{E}_f[\mathbb{E}_f[x|y]] = \mathbb{E}_f[x]$; that is, $\hat{x}$ is unbiased. Therefore, to preserve our central insight from section 5.1, we need to show that $\mathbb{E}_f[x] < \mathbb{E}_f[y]$ when the agent is risk averse.

Let the agent's cost-minimizing distribution $g(x,y)$ be a centered bivariate normal distribution with variance-covariance matrix $\Sigma_g = \begin{bmatrix} 1 & \rho \\ \rho & 1 \end{bmatrix}$. Then by definition of a conditional bivariate normal distribution, $\mathbb{E}_g[x|y] \equiv \rho y$. We can therefore use equation (65) to write $\mathbb{E}_f[y]$ as follows.

$$\begin{aligned} \mathbb{E}_f(y) &= \int\int y g(x,y) e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} dx dy \\ &= \int y e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} g(y) dy. \end{aligned} \tag{66}$$

Again using equation (65), we have

$$\begin{aligned} \mathbb{E}_f(x) &= \int\int x g(x,y) e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} dx dy \\ &= \int e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} g(y) \int x g(x|y) dx dy \\ &= \rho \int y e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} g(y) dy \\ &= \rho\, \mathbb{E}_f(y). \end{aligned} \tag{67}$$

Then for $\rho \in (0,1)$, $\mathbb{E}_f[x] < \mathbb{E}_f[y]$ if $\mathbb{E}_f(y) = \int y e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} g(y) dy > 0$.

Because $t(y) \equiv e^{U(\alpha^{-1}(\rho y+\lambda-\eta))-1-\nu} > 0$ and because $g(y)$ is symmetric around zero by our assumption that $g(x,y)$ is a centered multivariate normal, $\mathbb{E}_f[y] > 0$ if $t(y) > t(-y)$. We therefore want to show that

$$e^{U(\alpha^{-1}(\rho y+\lambda-\eta))} > e^{U(\alpha^{-1}(-\rho y+\lambda-\eta))} \tag{68}$$

for all y. The function $U(\cdot)$ is increasing in its argument and e^U is increasing in U, so it is sufficient to show that $\alpha^{-1}(\cdot)$ is monotonically increasing in its argument; that is, $(\alpha^{-1})'(s) > 0$. Recall that $\alpha(s) = \frac{1}{U'(s)} + s$ and assume that the agent has strictly monotonic utility so that there are no discontinuities in $\alpha(s)$. Then by the inverse function theorem,

$$(\alpha^{-1})'(s) = \frac{1}{\alpha'(s)} = \frac{U'(s)^2}{U'(s)^2 - U''(s)}. \tag{69}$$

By assumption, $U'(\cdot) > 0$ and $U''(\cdot) < 0$. Therefore $\alpha^{-1}(s) > 0$ for all s and $\mathbb{E}_f[x] = \rho\,\mathbb{E}_f[y] \leq \mathbb{E}_f[y]$ for $\rho \in [0,1]$.

B.2 Accounting for uncertain investments

In section 5.2 we used our tractable framework to study optimal measurement of uncertain investments. We showed that when information about future returns is perfectly reliable, fair value accounting is optimal, and when information about future returns is unreliable, optimal measurement depends on the correlation between current investments and future returns. We now show that this measurement insight is preserved when the agent is risk averse and when g is not restricted to be a normal.

Let x_1 be an investment made in the current period and let x_2 be the future return from that investment. Assume that the principal's objective is $B(\mathbf{x}) = b_2 x_2 - x_1$, where $b_2 > 1$. Let y_1 and y_2 be contractible evidence about x_1 and x_2. Assume that y_1 is perfectly reliable such that $\mathbb{E}_g[x_1|y_1] = y_1$. Equation (63) shows that $f(\mathbf{x}|\mathbf{y}) = g(\mathbf{x}|\mathbf{y})$ regardless of the agent's utility function or the distributional form of g. Then by Corollary 2, the optimal measurement and aggregation rules solving the principal's program are as follows.

$$\begin{aligned} \textit{Estimate of } x_1 \quad & \hat{x}_1 &=& \; y_1 \\ \textit{Estimate of } x_2 \quad & \hat{x}_2 &=& \; \mathbb{E}_g[x_2|y_1, y_2] \\ \textit{Net income} \quad & \pi(\hat{\mathbf{x}}) &=& \; b_2\hat{x}_2 - \hat{x}_1 \end{aligned} \tag{70}$$

Case 1: Future returns difficult to measure

We consider the case where evidence about x_2 is unreliable by assuming that y_2 is not informative about x_2. Then $\hat{x}_2 = \mathbb{E}_g[x_2|y_1, y_2] = \mathbb{E}_g[x_2|y_1] = \mathbb{E}_g[x_2|x_1]$. Then optimal measurement depends on the relationship between current investment and future returns. We highlight three subcases.

1. Suppose that there is no relationship between current and future returns: $\mathbb{E}_g[x_2|x_1] = \mathbb{E}[x_2] = 0$. Then $\hat{x}_2 = 0$ and net income is equal to $-\hat{x}_1$. This is akin to immediate expensing of highly uncertain investments such as R&D.

2. Consider an intermediate case in which $\mathbb{E}_g[x_2|x_1] = \frac{1}{b_2\,\mathbb{E}[x_1|y_1]}$. Then net income is equal to zero. Because the investment must go somewhere in a double entry system, we can interpret this as capitalization of the investment on the balance sheet.

3. Finally, assume that x_1 perfectly correlates to x_2; for simplicity assume that $E_g[x_2|x_1] = x_1$. Then $\hat{x_1} = \hat{x_2}$ and net income is given by $\hat{x}_1(b_2 - 1)$, akin to accruing revenue and expensing cost of goods sold for delivery of inventory sold on credit.

Case 2: Future returns easy to measure

Assume now that evidence about x_2 is perfectly reliable such that $\mathbb{E}_g[x_2|y_1, y_2] = y_2$. Then net income is given by $\pi(\hat{\mathbf{x}}) = b_2\hat{x}_2 - \hat{x}_1 = b_2 y_2 - y_1$. That is, the measurement process takes evidence as given and reports investment income at fair value.

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.

Baker, G. P. 'Incentive contracts and performance measurement.' Journal of Political Economy 100.3 (1992): 598–614.

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.

Banker, R. D., R. Huang and R. Natarajan. 'Incentive contracting and value relevance of earnings and cash flows.' Journal of Accounting Research 47.3 (2009): 647–678.

Beaver, W. H. 'Problems and paradoxes in the financial reporting of future events.' Accounting Horizons 5.4 (1991): 122–134.

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

Bloomfield, M., B. Gipper, J. D. Kepler and D. Tsui. 'Cost shielding in executive bonus plans.' Journal of Accounting and Economics (2021). In press.

Bonham, J. 'Shaping Incentives through Measurement and Contracts.' Unpublished paper, University of Chicago Booth School of Business, 2021. 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., E. Engel and A. Smith. 'An analysis of the relation between the stewardship and valuation roles of earnings.' Journal of Accounting Research 44.1 (2006): 53–83.

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

Butterworth, J. E., M. Gibbins and R. D. King. 'The structure of accounting theory: Some basic conceptual and methodological issues.' Research to support standard setting in financial accounting: A Canadian perspective 5 (1982): 1–65.

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.

Fama, E. F. and M. C. Jensen. 'Agency problems and residual claims.' The Journal of Law and Economics 26.2 (1983): 327–349.

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.

Gjesdal, F. 'Accounting for stewardship.' Journal of Accounting Research 19.1 (1981): 208–231.

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.

Hicks, J. R. Value and Capital. Oxford, U.K.: University Press, 1939.

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.

IJIRI, Y. The Foundations of Accounting Measurement: A Mathematical, Economic, and Behavioral Inquiry. Prentice-Hall, 1967.

——.Theory of Accounting Measurement. Sarasota, FL: American Accounting Association, 1975.

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.

——.'The role of accrual accounting in restricting dividends to shareholders.' European Accounting Review 7.4 (1998): 579–604.

MAY, G. O. Financial Accounting: A Distillation of Experience. New York: The Macmillan Company, 1943.

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

MOONITZ, M. Basic Postulates of Accounting. New York: American Institute of Certified Public Accountants, Accounting Research Study No. 01, 1961.

ORDELHEIDE, D. 'Kaufmännischer Periodengewinn als ökonomischer Gewinn.' Unternehmungserfolg. Springer, 1988. 275–302.

PATON, W. A. and A. C. LITTLETON. An introduction to corporate accounting standards. American Accounting Association Monograph No. 3, 1940.

PAUL, J. M. 'On the efficiency of stock-based compensation.' The Review of Financial Studies 5.3 (1992): 471–502.

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.

WATTS, R. L. and J. L. ZIMMERMAN. 'Positive accounting theory' (1986).

ZEFF, S. A. 'The objectives of financial reporting: A historical survey and analysis.' Accounting and Business Research 43.4 (2013): 262–327.