Contracting theory and accounting ☆

Richard A. Lambert*

The Wharton School, University of Pennsylvania, Philadelphia, PA 19104-6365, USA

Abstract

This paper reviews agency theory and its application to accounting issues. I discuss the formulation of models of incentive problems caused by moral hazard and adverse selection problems. I review theoretical research on the role of performance measures in compensation contracts, and I compare how information is aggregated for compensation purposes versus valuation purposes. I also review the literature on communication, including models where the revelation principle does not apply so that nontruthful reporting and earnings management can take place. The paper also discusses capital allocation within firms, including transfer pricing and cost allocation problems. © 2001 Elsevier Science B.V. All rights reserved.

JEL classification: D82; J33; L22; M41

Keywords: Incentives; Performance measurement; Aggregation; Resource allocation; Communication; Earnings management

1. Introduction

This paper reviews agency theory and its applications to accounting.[1] Agency theory has been one of the most important theoretical paradigms in

---

*Corresponding author. Tel.: +1-215-898-7782; fax: +1-215-573-5463.

E-mail address: lambert@wharton.upenn.edu (R.A. Lambert).

[1] See Arrow (1985), Baiman (1982, 1990), Hart and Holmstrom (1986), Kreps (1990), Milgrom and Roberts (1992), Pratt and Zeckhauser (1985), Prendergast (1999), and Indjejikian (1999) for prior reviews of aspects of agency theory.

accounting during the last 20 years. The primary feature of agency theory that has made it attractive to accounting researchers is that it allows us to explicitly incorporate conflicts of interest, incentive problems, and mechanisms for controlling incentive problems into our models. This is important because much of the motivation for accounting and auditing has to do with the control of incentive problems. For example, the reason we insist on having an "independent" auditor is that we do not believe we can trust managers to issue truthful reports on their own. Similarly, much of the motivation for focusing on objective and verifiable information and for conservatism in financial reporting lies with incentive problems. At the most fundamental level, agency theory is used in accounting research to address two questions: (i) how do features of information, accounting, and compensation systems affect (reduce or make worse) incentive problems and (ii) how does the existence of incentive problems affect the design and structure of information, accounting, and compensation systems?

While agency theory has generated insights into financial accounting and auditing issues, by far its largest contributions have been to managerial accounting. Accounting systems produce numerous measures of financial performance, including costs, revenues, and profits. Each of these financial measures of performance can be calculated at the "local" level or at higher levels, including the firm-wide level. The question of how to best measure performance is an important one because accounting and budgeting systems, performance measurement systems, transfer pricing systems, and decision support systems affect how people and organizations interact. Criticism continues to grow that traditional performance measures motivate dysfunctional behavior by causing managers to pay attention to the "wrong" things.

For example, many firms are beginning to place greater emphasis on nonfinancial measures such as quality, customer satisfaction, on time delivery, innovation measures, and on the attainment of strategic objectives.[2] Kaplan and Norton (1992, 1993) have developed the notion of a "balanced scorecard" to attempt to reflect the multi-dimensional nature of managerial performance and to capture value drivers in a more timely fashion than conventional accounting numbers. Consulting firms are developing and marketing alternative financial measures of performance such as economic value added, cash flow return on investment, shareholder value, etc. and claiming they provide "superior" measures of performance and better incentives in motivating managers to take the right actions. At the corporate level, the relative merits of stock price versus accounting numbers as measures of performance continue to be debated, and we have witnessed a tremendous increase in the use of stock-based compensation during the 1990s. Agency

---

[2] See Ittner et al. (1997) for empirical evidence on the use of nonfinancial measures in compensation contracts.

theory provides a framework for addressing these issues and rigorously examining the link between information systems, incentives, and behavior.

Agency theory has its roots in the information economics literature. As such, accounting and other information is placed into an explicit decision-making setting. The value of information is derived from the better decisions (and higher profits) that result from its use. Another important carryover from information economics is the idea that the most meaningful way to compare accounting/performance measurement systems is by comparing each system when it is used optimally. For example, in order for there to be a role for additional accounting information, it must be the case that the incentive problems being studied cannot be completely resolved via other means. This typically places restrictions on the type of "other" information that is assumed to be available in the model. It also forces the researcher to explicitly build uncertainty and measurement problems into the model.

The primary way agency theory distinguishes itself from "traditional" information economics is its belief that multi-person, incentive, asymmetric information, and/or coordination issues are important in understanding how organizations operate. To have an interesting multi-person model, agency researchers are careful to ensure that conflicts of interests are explicitly built into the analysis. That is, agency theory models are constructed based on the philosophy that it is important to examine incentive problems and their "resolution" in an economic setting in which the potential incentive problem actually exists.[3] Typical reasons for conflicts of interest include (i) effort aversion by the agent, (ii) the agent can divert resources for his private consumption or use, (iii) differential time horizons, e.g., the agent is less concerned about the future period effects of his current period actions because he does not expect to be with the firm or the agent is concerned about how his actions will affect others' assessments of his skill, which will affect
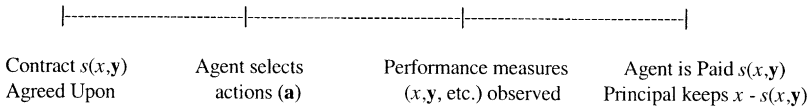
---

[3] In contrast, other literatures examine compensation and performance measures in settings where no conflict of interest is modeled. For example, prior to agency theory, many papers began with the assumption that compensation is an increasing function of say, divisional performance. These papers then examined the incentives of agents to distort their reported performance or the profitability of investing in their division (e.g., see Ronen and McKinney, 1970; Weitzman, 1976). However, in these models, there is generally no reason for making the agent's compensation an increasing function of his divisional performance. In these cases the incentive problem being studied can be trivially solved by simply paying the agent a fixed wage. That is, if the agent is paid a fixed wage he has no incentive to misreport his performance. Agency theory takes the perspective that if you want to analyze performance measurement systems, there are costs and benefits that interact. The cost of motivating misreporting through the use of a compensation system with a given property must be balanced against the benefit derived from choosing the compensation system to have that property in the first place. Unless the incentive problem that causes the compensation system to have that property to begin with is explicitly in the model, you cannot make such a tradeoff.

compensation in the future, or (iv) differential risk aversion on the part of the agent.

1.1. Setup of the basic agency model

In the simplest agency models, the organization is reduced to two people: the principal and the agent. The principal's roles are to supply capital, to bear risk, and to construct incentives, while the roles of the agent are to make decisions on the principal's behalf and to also bear risk (this is frequently of secondary concern). The principal can be thought of as a "representative shareholder" or the board of directors.[4] In more complicated agency models, there can be multiple principals and/or multiple agents. Some agents can even be both a principal and an agent, e.g., in a hierarchical firm a middle level manager might be the agent of managers above him and the principal to employees below him.

In order to more easily keep track of who knows what and when, it is often useful to construct a time line outlining the sequence of events in the model. In the "plain vanilla" principal–agent model, the sequence of events is as follows:

|---------------------------|------------------------------|------------------------------|

| Contract $s(x,\mathbf{y})$ | Agent selects | Performance measures | Agent is Paid $s(x,\mathbf{y})$ |
| Agreed Upon | actions ($\mathbf{a}$) | ($x,\mathbf{y}$, etc.) observed | Principal keeps $x - s(x,\mathbf{y})$ |

The principal selects a performance evaluation system which specifies the performance measures (or information signals) upon which the agent's compensation will be based and the form of the function that links the performance measures to the agent's compensation. Let s denote the compensation function, and $\mathbf{y}$ the vector of performance measures to be used in the contract. Based on this contract, the agent selects a vector of actions, $\mathbf{a}$, which could include operating decisions, financing decisions, or investment decisions. These decisions, along with other exogenous factors (generally modeled as random variables) influence the realizations of the performance measures, as well as the "outcome" of the firm, which we denote as x.

We will assume the outcome is measured in monetary terms, although in some contexts such as health care choices or government policy choices, the outcome might be better thought of as being nonmonetary. In a single-period model, the monetary outcome is well defined; it represents the end-of-period cash flow or the liquidating dividend of the firm gross of the compensation paid to the agent. For now, we will assume that the outcome x is observable and can

---

[4] In other contexts the principal and the agent could be (i) bondholders versus shareholders, (ii) regulator versus regulatee, (iii) citizens versus government policy makers, (iv) doctor versus patient, (v) two separate firms, etc.

be contracted on. This assumption will be relaxed later. After the performance measures are jointly observed, the agent is paid according to the terms of the contract. Note that this formulation implicitly assumes that the property rights to the outcome belong with the principal. A few papers consider the opposite situation in which the agent has the property rights to the outcome by allowing him to keep any "unreported income".

The "plain vanilla" version of the agency model has been extended in a number of ways. For example, as mentioned above, the outcome might not be observable. In this case there is potentially a role for information that helps estimate the outcome. Considerable effort in accounting research has also been directed at modeling different mechanisms by which the information signals, $\mathbf{y}$, are produced. The simplest case is that they are simply "generated" by the actions and "automatically" observed by the parties. Other papers have modeled the situation where the principal observes some information at the end of the period and then decides whether to conduct an investigation to obtain more information (e.g., a variance investigation).[5] Another possibility is that the information is generated by a report made by the agent. In this case, there may be moral hazard problems on the agent reporting truthfully. The information signal might also be generated via a third party such as an auditor. In this case, incentives problems with the auditor (e.g., independence or how intensively does he audit and does he report his findings honestly) can be modeled and analyzed.[6] Finally, the performance measure may come from the security market's process of aggregating information into stock prices. Again, issues regarding what information is available to investors, and how this information is affected by operating and reporting decisions by the agent can be modeled and analyzed.

Agency papers have also extended the basic model by allowing the agent and/or the principal to obtain information prior to the agent selecting his action. This information could relate to the productivity of different operating actions, the general "favorableness" of the environment, or information about the employee's type (e.g., his skill or his risk aversion). The pre-decision information could be received before the contract is signed or between the time the contract is signed and the time the agent selects his actions. In these papers, communication of the agent's information via participative budgeting can be studied. Agency papers have also extended the basic model to include multi-periods (either where a single-period model is repeated over time or where there are explicit interdependencies between the periods).

Finally, papers have modeled issues that arise when there are multiple agents in the firm. This enables us to examine the role of encouraging/discouraging

---

[5] See Baiman and Demski (1980), Lambert (1985), Young (1986), and Dye (1986) for more analysis.

[6] See Antle (1982), Baiman et al. (1987), and Baiman et al. (1991) for examples.

competition among agents, and the use of relative performance to compare the performance of agents. With multi-agent models we can also study the interaction between management accounting and organizational structure, including hierarchies, job design, and task allocation. Multi-agent models are also necessary to studying the role of incentive problems in allocating resources (and costs) among agents, and analyzing transfer pricing between subunits.

1.2. Organization of the paper

In the next section, I discuss single-period, single-action agency models in which the incentive problem arises because the agent's actions are unobservable to the principal. These types of incentive problems are referred to as moral hazard or hidden action problems. I describe the features the models must possess in order for a genuine incentive problem to exist that cannot be costlessly resolved. I then discuss the role of performance measures in reducing the magnitude of the agency problem. The key characteristic here is the informativeness of the performance measure about the agent's action. The informativeness of a performance measure is a function of its sensitivity to the agent's actions and its noisiness. I discuss the implications of these models for the shape of the optimal contract, the conditions where performance measures are combined in a linear fashion (which is how accounting systems aggregate line items), and the ideas of responsibility accounting and the controllability principle commonly discussed in managerial accounting textbooks.

In Section 3, I continue to analyze hidden action models, but in models where the agent is responsible for multiple actions. In this section, I discuss the Linear contracting, Exponential utility function, Normal distribution (LEN) framework for formulating agency problems. In a multi-action model, the emphasis shifts from that of motivating the intensity of the agent's effort to the allocation of his effort. Accordingly, the congruity of a performance measure (or how it contributes to constructing an overall performance measure that is congruent) becomes important. I discuss the application of the results to accounting "window dressing" and earnings management, to incomplete or myopic measures of performance, to the role of nonfinancial measures of performance, divisional versus firm-wide performance, the valuation versus stewardship uses of information, and stock price versus accounting numbers in compensation contracts.

In Section 4, I focus on agency problems caused by the agent possessing superior information about a parameter that affects the outcome-generating process or perhaps about the outcome itself. In these models, accounting systems are used to communicate information within the organization, to coordinate actions across parties, as well as to evaluate the actions that have been taken and the outcome that has occurred. A new role for accounting systems is to reduce the "information rent" that the agent is able to extract

based on his information advantage. I discuss the application of these results to issues of participative budgeting and target setting (including the creation of organization slack), to the "confirmatory" role of accounting numbers, hurdle rates for allocating capital, transfer pricing, and cost allocation.

Section 5 discusses communication, earnings management, and the revelation principle. In particular, I describe the qualities a model must possess to circumvent the revelation principle so that earnings management issues can be addressed. In Section 6, I briefly discuss multiple period agency models. Multi-period models are essential for earnings and cash flows to be different and for accruals to have a substantive role. I discuss multi-periods models in motivating long-term investment decisions, the use of cash flow versus accrual accounting versus residual income measures of performance, and the role of depreciation policies. The final section outlines some suggestions for future research.

2. Single-action agency models

In words, we express the principal's problem as a constrained maximization problem in which he chooses the compensation function (its form and the variables it is based on) to[7]

maximize the principal's expected utility

subject to   agent's acceptable utility constraint,

agent's incentive compatibility constraints.

The principal's utility is defined over the "net" proceeds generated by the firm, e.g., the outcome x minus the agent's compensation, s. Let $G[x - s]$ denote the principal's utility function. The principal is assumed to prefer more money to less, $G' > 0$, and be risk averse or risk neutral, $G'' \leqslant 0$. For a risk neutral principal, his expected utility is simply the expected net profits of the firm. For risk averse principals, higher moments of the distribution of net profits are also important. As I discuss below, it is common to assume the principal is risk neutral.

The net profits to the principal are influenced by the compensation function in two ways. First, there is a direct effect, because each dollar paid to the agent as compensation is a dollar less for the principal. Second, there is an incentive effect, because the structure of the compensation function will affect the actions

---

[7] In some models, it may also be important to include a floor on the payments made to the agent, e.g., the agent's payment cannot be negative, which would imply the agent is paying the principal. Similarly, in some models it may be useful to explicitly include a constraint which specifies a maximum payment. For example, the agent's payment might be constrained to be less than the outcome, x.

selected by the agent, which will affect the probability distribution of the gross outcome, x. The outcome and the performance measures are also affected by other factors that are treated as exogenous to the model. We model these by assuming the outcome and performance measures are random variables whose distributions are affected by the agent's actions. Let $f(x, \mathbf{y}|\mathbf{a})$ denote the joint probability density of the outcome x and the performance measures $\mathbf{y}$ given the agent's actions.[8] For the most part, I will assume the variables x and $\mathbf{y}$ are continuous random variables; however, at times I will discuss models where they are discrete. Initially we will assume that the principal and agent have homogenous beliefs about the distribution $f(x, \mathbf{y}|\mathbf{a})$. In later sections I will consider situations in which one party has superior information.

In choosing a compensation function, the principal must ensure that it is attractive enough to offer the agent an "acceptable" level of expected utility. This is typically modeled as requiring the agent's expected utility from the contract offered by the principal to meet some exogenously specified minimal acceptable level. This minimal level is often interpreted as the expected utility of the agent in his next best employment opportunity, or his reservation level of utility. This interpretation suggests that the principal has all the "power" in the relationship; he can hold the agent to this minimal acceptable level, while he keeps the excess.[9] However, an alternative interpretation of the agency formulation is that it is merely trying to identify Pareto Optimal outcomes. That is, we can view the minimal acceptable level of utility for the agent as already reflecting the bargaining power of the agent. By varying the minimal acceptable level of utility for the agent, we can sweep out the Pareto frontier of achievable combinations of expected utilities of the two parties.

The second set of constraints, termed incentive compatibility constraints, represents the link between the contract chosen and the actions selected. Given the contract offered, the agent will choose the actions (and messages if there is a communication dimension to the model) that maximize his expected utility. Including the incentive compatibility constraints allows us to model the agency problem *as if* the principal is selecting both the contract and the actions, but the principal is constrained to choose a (contract, action) combination that is incentive compatible for the agent. As I discuss in a later section, researchers have had difficulty modeling the incentive compatibility constraints; a number of different mathematical approaches have been used.

---

[8] Since the actions $\mathbf{a}$ are not random variables (at least not in the simplest models), it is not literally correct to refer to the distribution of the outcome as being conditional upon the actions $\mathbf{a}$. A better way to phrase it is that the probability distribution is parameterized by the actions, $\mathbf{a}$.

[9] Some papers explicitly assume the opposite power structure, that is, they assume that competition among principals will drive their levels of expected profits to zero. In this case, the agent is assumed to capture all the "excess".

The agent's utility function is defined over both his monetary compensation, s, and the actions be selects, $\mathbf{a}$. In most of the agency literature, the agent's utility function is assumed to be additively separable into compensation and action components, $H(s, \mathbf{a}) = U(s) - V(\mathbf{a})$. However, some models assume multiplicative separability, $H(s, \mathbf{a}) = U(s)V(\mathbf{a})$. The most common interpretation of the nonmonetary portion of the utility function is that the agent's actions represent the effort levels he puts into various activities. More effort is assumed to increase the expected outcome, but to be personally costly for the agent. In other models, the nonpecuniary return associated with the actions is interpreted as power, prestige, or resources diverted for personal use or consumption.

Consistent with most agency theory papers, I have modeled the agent's monetary utility as being defined solely over the compensation he receives from the compensation contract. This places a large burden on the compensation contract because it is the only source of incentives for the agent. In reality, there are numerous other forms of incentives, including monetary incentives from other sources (i.e., the labor market or takeover market) and nonmonetary incentives (satisfaction, embarrassment, promotion, jail time, etc.). This assumption also implicitly assumes that the agent has no other wealth, or that the principal is somehow able to contract over the agent's entire wealth. This allows the principal to decide what risks the agent bears and how the agent's consumption is allocated over time (in multi-period models). In general, the choices the principal would make for the agent along these lines would not be the ones the agent would make himself.[10] In particular, if the agent has access to insurance markets or capital markets, he may choose to offset or hedge some of the risk the principal desires to impose on him, or borrow against future earnings, etc. To the extent these issues are thought to have an important impact on the incentive problem being examined, the model should incorporate outside wealth for the agent, and it should be clear about what opportunities the agent has to re-allocate this wealth in response to the contract offered by the principal.

2.1. First-best solution

As a benchmark for comparison, agency theory papers generally first compute the solution to the agency problem assuming away the incentive problems. In this so-called "first-best" solution, the actions are chosen cooperatively with all parties' interests in mind and all reports are issued truthfully. Mathematically, we can express the first-best solution as the

---

[10] See Fischer (1999) for analysis of a model which compares compensation risk management decisions by the firm versus by the manager.

solution to the following problem:

$$\underset{s(x,\mathbf{y}),\mathbf{a}}{\text{maximize}} \iint G[x - s(x,\mathbf{y})]f(x,\mathbf{y}|\mathbf{a})\,\mathrm{d}x\,\mathrm{d}\mathbf{y}$$

$$\text{subject to} \iint U[s(x,\mathbf{y})]f(x,\mathbf{y}|\mathbf{a})\,\mathrm{d}x\,\mathrm{d}\mathbf{y} - V(\mathbf{a}) \geqslant \underline{H}.$$

That is, in the first-best solution, we choose the contract and the actions to maximize the principal's expected utility subject to meeting the agent's acceptable level of utility, $\underline{H}$. Note that there is no incentive compatibility constraint present because the actions are not chosen "selfishly" by the agent; they are chosen cooperatively.

Letting λ be the Lagrange multiplier on the acceptable utility for the agent constraint, we can express the first-best solution as the solution to the problem

$$\underset{s(x,\mathbf{y}),\mathbf{a}}{\max} \iint G[x - s(x,\mathbf{y})]f(x,\mathbf{y}|\mathbf{a})\,\mathrm{d}x\,\mathrm{d}\mathbf{y}$$
$$+ \lambda \left\{ \iint U[s(x,\mathbf{y})]f(x,\mathbf{y}|\mathbf{a})\,\mathrm{d}x\,\mathrm{d}\mathbf{y} - V(\mathbf{a}) - \underline{H} \right\}. \tag{1}$$

With this formulation, we can see that the agency problem can be thought of as maximizing a weighted combination of the expected utilities of the principal and agent. By varying λ (or equivalently by varying $\underline{H}$), we can sweep out the entire Pareto frontier.[11]

The optimal contract is derived by differentiating the objective function with respect to s for each possible $(x,\mathbf{y})$ realization. The first-order condition is

$$-G'[x - s(x,\mathbf{y})] + \lambda U'[s(x,\mathbf{y})] = 0$$

which can be re-expressed as

$$\frac{G'[x - s(x,\mathbf{y})]}{U'[s(x,\mathbf{y})]} = \lambda. \tag{2}$$

This equation shows that the agent's compensation is set to make the ratio of the marginal utilities of the principal and the agent a constant across all $(x,\mathbf{y})$ realizations. This is referred to as the optimal risk sharing condition, and dates back to Wilson's (1968) work on syndicate theory.

Eq. (2) implies that the contract depends only on the outcome x and that the performance measures $\mathbf{y}$ are not used in the contract.[12] That is, as long as the

---

[11] Technically, these are only equivalent if the shape of the Pareto frontier is concave. If it is not concave, there are gains to randomizing over (contract, action) pairs in order to make the frontier concave. For the range of the frontier where randomization is optimal, the same welfare weight (λ) will apply for the agent. See Fellingham et al. (1984) and Arnott and Stiglitz (1988) for a discussion of ex ante randomization in agency settings.

[12] When at least one of the principal and the agent is risk averse, this follows directly from Eq. (2). When both parties are risk neutral, there are an infinite number of contracts which yield the same levels of expected utilities. Any contract which includes $\mathbf{y}$ in a nontrivial way can be weakly dominated by the one that is based solely on x.

outcome x is observable, there is no role for additional performance measures. The outcome x is the only "real" risk in the model, so it is the only one that is relevant for risk sharing purposes. Since there are (by definition) no incentive problems in the first-best solution, there is no other role for other performance measures.

Note that if the principal is risk neutral and the agent is risk averse, the optimal contract satisfies

$$\frac{1}{U'[s(x, \mathbf{y})]} = \lambda.$$

Since the right-hand side of this equation is a constant, the left-hand side must also be a constant. This implies that the optimal risk sharing contract pays the agent a constant, $s(x, \mathbf{y}) = k$. That is, a risk neutral principal does not mind bearing all the risk, while a risk averse agent prefers not to bear any risk, ceteris paribus. Therefore, the optimal contract completely shields the agent from any risk. Similarly, if the agent is risk neutral and the principal is risk averse, the optimal risk sharing contract is for the agent to bear all the risk; the optimal contract is $s(x, \mathbf{y}) = x - k$. When both parties are risk averse, it is optimal for each party to bear some of the risk. The shape of the optimal risk sharing contract depends on the specific forms of the two parties' utility functions. An interesting special case is where each party has negative exponential utility. In this case, the optimal risk sharing contract is linear in the outcome, x, and the slope coefficient (which is the amount of risk borne by the agent) is proportional to the risk tolerance of the agent relative to that of the principal. See Wilson (1968) for additional discussion.

2.2. Situations where the first-best solution can be achieved

We now turn to models in which the principal must also consider the incentive aspect of the compensation scheme he offers. That is, we now assume that the agent will select the actions that are in his own best interests given the compensation scheme offered by the principal. We begin by noting a number of special situations where the first-best solution can still be achieved. In these cases, the principal can construct a compensation scheme that both shares risk optimally and simultaneously gives the agent incentive to select the action that was chosen in the first-best solution. Note that it is not sufficient that the principal can design a contract to induce the agent to select the first-best actions. The contract must also do this without imposing more risk than the pure risk sharing contract offers.

The first special case we consider is when the agent is risk neutral. In this case, the optimal risk sharing contract causes the agent to bear all the risk. Since the principal receives a constant payment, the first-best objective function in Eq. (1) is equivalent to one in which the actions are chosen to maximize the

agent's expected utility. That is, the first-best action is the one that maximizes the agent's expected utility given the optimal risk sharing contract. Obviously, the agent's private incentives are to select this same action. Therefore, if we "sell" the firm to the agent, he internalizes the incentive problem. With a risk neutral agent, selling the firm is also optimal from a risk sharing perspective. Note that in order for this to work, it has to be possible to transfer the property rights to the outcome to the agent.[13] Moreover, the agent must have sufficient wealth to be able to absorb any loses that may occur ex post.

Next, we examine situations in which the first-best solution can be achieved even if the agent is risk averse. The most obvious case is where the agent's actions are observable. In this case the principal can offer a "compound" contract in which (i) the agent is paid according to the terms of the first-best contract if the principal observes the agent has selected the first-best actions and (ii) the agent is penalized substantially if the principal observes any actions other than the first best. Since any deviation from the first-best action will be detected with certainty, the agent will select the first-best actions to avoid the penalty. This type of contract is known as a "forcing" contract because it forces the agent to select the action the principal specifies. Since the compensation contract also shares risk optimally given that the first-best action is taken, the first-best solution is achieved.

The first-best solution can also be achieved if there is no uncertainty in the outcome distribution. In this case, the principal can infer from the outcome that occurs whether the first-best action has been selected. He can therefore penalize any deviations from first best to ensure the first-best action is chosen. The first-best solution can also be achieved even if there is uncertainty if the "state of nature" and the outcome are both observable. Similar to the prior cases, the principal can invert the outcome function and infer whether the first-best action was selected. The contract can again offer the agent the optimal risk sharing terms if the "correct" actions were taken and a substantial penalty otherwise.

Finally, the first-best solution can sometimes be achieved if the outcome distribution exhibits "moving support". This means that the set of possible outcomes changes with the actions selected. For example, suppose the agent is responsible for deciding how "hard" to work, and that if he selects an effort level of a, the outcome is uniformly distributed between $[a - c, a + c]$. If the first-best action is a^*, then any realized outcome that falls within the range of $[a^* - c, a^* + c]$ is *consistent* with the first-best action being taken, but does not *imply* the first-best action was taken. While the agent could put in slightly less

---

[13] If the principal and the agent are both risk neutral, it is not necessary to be able to "sell" the firm to the agent to be able to achieve first best. All that is necessary is that there be noisy signal that is statistically related to the agent's effort available for contracting (and certain regularity conditions are met).

effort than $a*$ and still have a good chance of having the outcome fall within this range, there is also some chance that the realized outcome will be below $a*-c$. If this occurs, the principal knows for certain that the agent did not select the desired action. If large enough penalties are available to the principal to penalize these outcomes, the principal can ensure the agent will select the first-best action. That is, if the principal offers the agent a contract that pays him identically to the optimal risk sharing contract if $x \in [a*-c, a*+c]$, and some extremely small amount (or even requires the agent to pay compensation to the principal) if $x < a*-c$, the agent will select the first-best action. Moreover, as long as the agent selects the first-best action, the contract offers him optimal risk sharing.

These examples demonstrate that if we only look at contract payments that occur, it may appear there are little incentives present. That is, in the latter examples, the incentives in the contract are so powerful that, in equilibrium, shirking never occurs and penalties are never levied. A researcher who only observes the realized outcomes and payments might incorrectly infer from this that there are no incentives in the contract. These examples also illustrate the importance of the information role of performance measures in contracting. The examples are extreme in that there is always some chance we will observe a signal that unambiguously tells us the agent deviated from the desired action. More generally, we would expect information signals to be imperfect indicators of the agent's actions. We turn to these situations next.

2.3. Second-best solution—only the outcome (x) is observable

In this section, we begin to analyze models in which contracting on the outcome alone leads to a welfare loss relative to first best.[14] This is a necessary condition for additional information to have value in an agency setting. We begin with the simplest case where the agent is responsible for only a single-dimensional action: how much effort to supply, $a \in A$, where A is the set of feasible actions. We will assume effort is a continuous variable, and that the outcome is also a continuous random variable. Let $f(x|a)$ denote the probability density of the outcome for a given level of effort. More effort is assumed to increase the expected output; specifically, higher effort levels shift the probability distribution of the outcome to the right in the sense of first-order stochastic dominance. We rule out the "moving support" scenario by assuming that no outcome realization can be used to unambiguously confirm or reject any action in the feasible set, A. That is, if $f(x|a) > 0$ for some a, then it is also positive for all other actions in the feasible set. Finally, we assume that higher levels of effort are more personally costly to the agent, i.e., he has disutility for effort. In particular, we assume $V'(a) > 0$ and $V''(a) > 0$.

---

[14] The discussion here closely follows the analysis in Holmstrom (1979).

We can write the principal's problem as

$$\underset{s(x),a}{\text{maximize}} \int G[x - s(x)] f(x|a)\, dx \tag{3}$$

$$\text{subject to} \int U[s(x)] f(x|a)\, dx - V(a) \geqslant \underline{H} \tag{3a}$$

$$a \text{ maximizes} \int U[s(x)] f(x|a)\, dx - V(a). \tag{3b}$$

As in the first-best solution, we maximize the principal's expected utility subject to offering the agent an acceptable level of expected utility.[15] However, we now add the incentive compatibility constraint (Eq. (3b)) that states that the action chosen is the one that maximizes the agent's expected utility given the contract offered by the principal.

Unfortunately, the last constraint is not very tractable in its present form. Agency theorists have tried to recast it in a more tractable form by replacing it with the agent's first-order condition on effort.[16] Assuming the optimal effort is in the interior of the action set, the agent's optimal effort choice will be one at which the derivative of his expected utility with respect to his effort is equal to zero. The agent's first-order condition on effort is

$$\int U[s(x)] f_a(x|a)\, dx - V'(a) = 0. \tag{4}$$

Agency theory researchers substitute this first-order condition on effort, which is a "simple" equality constraint, for the incentive compatibility constraint.[17]

---

[15] If there are no constraints on the minimal payment that can be made to the agent or if the agent's utility function is unbounded below as his payment approaches any minimal allowable payment, then the principal can hold the agent to the minimal acceptable utility level in expectation.

[16] An alternative way to develop a tractable formulation of the agent's action choice is to assume the agent's choice set is discrete, that is, he selects from a finite number of possible actions. In this case, the agent's incentive compatibility constraint can be represented with a set of inequality constraints. For each action in the feasible set, the agent's expected utility under that action must be less than or equal to his expected utility under the action the principal wishes him to choose given the contract offered. Some of these constraints will be binding, which means that the agent is indifferent between those two actions given the contract. When the agent is indifferent over a set of actions, he is presumed to choose the action most preferred by the principal. See Grossman and Hart (1983) for more discussion and analysis of discrete action agency models.

[17] The problem with this approach is that the first-order condition will be satisfied for all actions that are local minima or maxima for the agent, not just the action that is the global maximum for the agent. When researchers use the first-order condition to represent the incentive compatibility constraint, they risk making the mistake of pairing a contract with one of the other local maxima or minima. The researcher thinks he has calculated a (contract, action) equilibrium, when in fact the agent would never select that action given that contract. Researchers have attempted to derive conditions where they can rule this out (see Grossman and Hart, 1983; Rogerson, 1985; Jewitt, 1988). Essentially, these papers develop conditions that ensure that the agent's expected utility is a

Letting λ be the Lagrange multiplier on the acceptable utility constraint and μ be the Lagrange multiplier on the agent's first-order condition on effort, the principal's problem becomes

$$\underset{s(x),a}{\text{maximize}} \quad \int G[x - s(x)]f(x|a)\,\mathrm{d}x + \lambda\left\{ \int U[s(x)]f(x|a)\,\mathrm{d}x - V(a) - \underline{H} \right\}$$
$$+ \mu\left\{ \int U[s(x)]f_a(x|a)\,\mathrm{d}x - V'(a) \right\}. \tag{5}$$

We characterize the optimal contract by taking the derivative of this problem with respect to s for each value of x. The resulting first-order condition for the optimal contract is

$$-G'[x - s(x)]f(x|a) + \lambda U'[s(x)]f(x|a) + \mu U'[s(x)]f_a(x|a) = 0,$$

which we can re-arrange as

$$\frac{G'[x - s(x)]}{U'[s(x)]} = \lambda + \mu \frac{f_a(x|a)}{f(x|a)}. \tag{6}$$

Note that if $\mu = 0$, this reduces to the optimal risk sharing contract in first-best solution. Therefore, the test of whether the first-best solution is achievable is equivalent to testing whether μ, the Lagrange multiplier on the incentive compatibility constraint, is nonzero. Holmstrom (1979) shows that $\mu > 0$ as long as the principal wants to motivate more than the lowest possible level of effort in A. The easiest way to see this is to consider the special case where the principal is risk neutral. In this case, if $\mu = 0$, the agent's compensation is a fixed wage, which makes his compensation independent of the outcome and independent of his effort. Under such a contract, the agent has no incentive to work hard, so he provides the minimal possible effort. Therefore, $\mu = 0$ cannot be the optimal solution.

With $\mu > 0$, the optimal contract deviates from optimal risk sharing depending on the sign and the magnitude of the term $f_a(x|a)/f(x|a)$. Milgrom (1981) shows that this term can be interpreted in terms of classical statistical inference. That is, suppose we are attempting to estimate a parameter of a probability distribution using maximum likelihood estimation methods. Specifically, suppose we observe a sample outcome x and the probability distribution of x is $f(x|a)$, where a is the parameter to be estimated. The maximum likelihood estimate is constructed by first taking the log of the likelihood function, then taking the derivative of this with respect to the parameter to be estimated, and finally setting the derivative equal to zero. Performing these calculation yields $\partial \log[f(x|a)]/\partial a = (1/f(x|a))f_a(x|a)$, which is

---

strictly concave function of his effort. In this case, the first-order condition will only be satisfied by one action, and that action will be the global maximizer for the agent (assuming it is interior). These conditions are generally very strong, but note that they are merely sufficient (but not necessary) for the first-order condition approach to be valid.

the same expression as appears in the characterization of the contract in Eq. (6). From a statistical perspective, we can think of the principal as trying to use the outcome x to try to infer whether the correct level of effort has been taken.[18] The principal rewards those outcomes that indicate the agent worked hard (if $f_a(x|a)$ is positive, this outcome is more likely to happen if the agent works harder) and penalizes those outcomes that indicate the agent did not work hard (i.e., harder work makes less likely those outcomes for which $f_a(x|a)$ is negative).

The optimal contract trades off the benefits of imposing risk in order to give the agent incentives with the cost of imposing risk on a risk averse agent. That is, if the principal offers the optimal risk sharing contract, the agent does not have enough incentive to provide a high enough level of effort. However, imposing risk on the agent lowers his expected utility ceteris paribus, so the principal must raise the agent's expected compensation to meet the agent's acceptable utility constraint.

Eq. (6) indicates that the shape of the contract depends on the functional form of the term $f_a(x|a)/f(x|a)$ as well as the shape of the principal's and agent's utility functions. As we show below, it is easy to construct examples where the optimal contract is linear, convex, or concave. This is both a blessing and a curse of agency theory. It is a blessing in the sense that the framework can be used to explain a wide variety of contract shapes. The curse is that many of the results depend on parameters such as the agent's degree of risk aversion, which is unlikely to be observable to a researcher.

Unless additional structure is imposed, it is difficult to prove even such basic properties as the optimal contract being increasing in the firm's outcome.[19] A sufficient condition to ensure that the contract is increasing in the outcome is if $f_a(x|a)/f(x|a)$ is increasing x. This is a stronger condition than assuming that the mean of the outcome distribution is increasing in the agent's effort. It also requires that higher effort makes all higher values of x relatively more likely than it makes lower values of x. We can also ensure monotonicity in the contract by enriching the set of actions available to the agent (see Verrecchia, 1986). In particular, suppose the agent privately observes the outcome first and has the ability to "destroy" some of the outcome before it is observed by the principal. If the compensation contract has any ranges where it is decreasing, the agent will destroy output to the point where his compensation no longer

---

[18] The statistical analogy is not exact, because the effort level is not a random variable. In fact, the principal knows exactly what the effort will be in response to the contract offered. Unfortunately, the principal cannot observe the effort level, so he cannot verify his conjecture in a way that can be used in the contract.

[19] See Grossman and Hart (1983) and Verrecchia (1986) for examples where the likelihood function $f_a(x|a)/f(x|a)$ is not monotonically increasing in x, and as a result, the optimal compensation contract has regions where it decreases.

decreases. Since outcomes in the range where compensation is decreasing would never be observed, the principal can duplicate the contract by making it flat over these ranges. This gives the agent the same "productive" incentives as the contract that contains decreasing ranges, and it removes the agent's incentive to destroy output. Therefore, the principal strictly prefers the nonmonotonic contract to the one that contains decreasing ranges.

As we discuss in more depth in Section 2.5, for many common probability distributions, the likelihood ratio $f_a(x|a)/f(x|a)$ is monotonic and *linear* in the outcome x. Eq. (6) demonstrates that a linear likelihood function does not imply that the optimal compensation contract is linear in the outcome. The contract form also depends on the shape of the principal's and agent's utility functions. To illustrate this, suppose the principal is risk neutral and the agent's utility function is a member of the hyperbolic absolute risk aversion (HARA) class of utility functions: $U(s) = [1/(1-\gamma)](\delta_0 + \delta_1 s)^{1-\gamma}$. This is a rich class of utility functions that is widely used in economics research. For example, members of the HARA class include the power utility functions and the logarithmic utility (as γ approaches 1). The HARA class can also be transformed to yield the negative exponential utility functions. The parameter $\gamma > 0$ is a measure of the agent's risk aversion.

For this class of utility functions, we have $1/U'(s) = (1/\delta_1)(\delta_0 + \delta_1 s)^{\gamma}$, so the optimal contract satisfies

$$s(x) = -\frac{\delta_0}{\delta_1} + \delta_1^{(1-\gamma)/\gamma}\left[\lambda + \mu \frac{f_a(x|a)}{f(x|a)}\right]^{1/\gamma}. \tag{7}$$

Assuming $\mu > 0$ and $f_a(x|a)/f(x|a)$ linear in x, Eq. (7) indicates that the optimal compensation function is a concave function of the outcome x if $\gamma > 1$. Similarly, the contract is linear in x if $\gamma = 1$ (i.e., the logarithmic utility function), and the contract is convex in x if $0 < \gamma < 1$. Therefore, the contract shape depends in part on how risk averse the agent is.[20]

Technically, the first-order condition given in Eq. (6) or (7) applies only if the payment is in the interior of the set of feasible payments. In many instances, payments must be bounded below. For example, limited liability or wealth on the part of the agent may prevent the principal from "penalizing" the agent too severely for bad outcomes. Similarly, note that the left-hand side of Eq. (6) is the ratio of marginal utilities, which is a positive number. Therefore, the right-hand side of Eq. (6) must also be positive. However, the term $f_a(x|a)/f(x|a)$ has an expected value of zero, so it can take on negative values. There is no guarantee that the parameters λ and μ have magnitudes that ensure the right-hand side of Eq. (6) will remain positive. To illustrate this, if x is normally distributed, $f_a(x|a)/f(x|a)$ is unbounded below, so as long as $\mu \neq 0$, there is always a range of outcomes for which the right-hand side of Eq. (6) will be

---

[20] See Hemmer et al. (1999) for further analysis on the shape of the contract.

negative. Therefore, there is no solution to the agency problem unless a lower bound on the payments is imposed, and there will always be a range of outcomes where this lower bound is binding.[21] As a result, optimal contracts will often be nondifferentiable, e.g., they may be piecewise linear. I will return to the issue of contract shape in later sections.

2.4. When are additional performance measures valuable?

In the previous section, we documented situations where a welfare loss occurs relative to the first-best solution. This means that additional performance measures can potentially increase the expected utilities of the principal and the agent if they can be used to increase the incentives or improve the risk sharing of the contract. To analyze this, consider a straightforward modification of the model in the previous section to make both the outcome x and an additional performance measure y observable and available for contracting.

Analogous to Eq. (6), the first-order condition on the optimal sharing rule satisfies

$$\frac{G'[x - s(x, y)]}{U'[s(x, y)]} = \lambda + \mu \frac{f_a(x, y|a)}{f(x, y|a)}. \tag{8}$$

As in the previous section, it can be shown that $\mu > 0$. This means that the optimal contract depends on the performance measure y if and only if the term $f_a(x, y|a)/f(x, y|a)$ depends on y.

Holmstrom (1979) shows that this condition has a statistical analogy in terms of sufficient statistics. That is, if we view the action a as a random parameter we are trying to estimate, the term $f_a(x, y|a)/f(x, y|a)$ depends on y unless it is the case that x is a sufficient statistic for x and y with respect to a.[22] For example, if we can write $x = a + \tilde{e}_1$ and $y = x + \tilde{e}_2 = a + \tilde{e}_1 + \tilde{e}_2$ where $\tilde{e}_1$ and $\tilde{e}_2$ are independent random variables, then even though y is "informative" about the agent's action by itself, it does not add any information about a that is not already conveyed by x. Therefore, there is no reason to add y to the contract if x is also available.

Holmstrom's informativeness condition suggests that contracts will be rich and based on many variables. While it is not surprising that a variable is not valuable if the other available variables are sufficient for it, it is more surprising that a variable is valuable as long as the other available variables are not

---

[21] This issue also raises questions regarding the existence of a solution to the principal's problem. See Mirrlees (1974) for analysis.

[22] In general, the density function of x and y can be expressed as $f(x, y|a) = h(x|a)g(y|x, a)$, and the likelihood ratio term in the optimal contract is $f_a(x, y|a)/f(x, y|a) = h_a(x|a)/h(x|a) + g_a(y|x, a)/g(y|x, a)$. When x is a sufficient statistic for y the density function reduces to $f(x, y|a) = h(x|a)g(y|x)$, and the dependence of f_a/f on y goes away.

sufficient for it. In particular, it seems plausible that a variable could be "slightly" informative, but be so noisy that its use would add too much risk into the contract. The fact that this is not the case must relate to "how" the variable is used in the contract. We turn to this subject next.

2.5. Aggregation of performance measures

Holmstrom's informativeness condition tells us when a new variable has nonzero value, but it does not indicate the factors that determine how *much* value it has, or how it is used in the contract. In this section we discuss in more detail the functional form of the contract and how signals are aggregated. One problem with interpreting the results of agency theory comes from the fact that both the form of the compensation function and the method of aggregating the signals are determined jointly. We end up with a mapping from the basic signals into a compensation number, and it can be difficult to separately interpret the information aggregation process from the compensation function form. Since accountants are typically more responsible for the performance measurement process and less for the compensation scheme construction, it is of great interest to be able to separate the two. Accounting systems generally aggregate signals in a linear fashion, so it is important to understand when such a linear aggregation is optimal. Note that linear aggregation of signals does not mean that the contract is linear.

For convenience I will discuss this issue for the case where the principal is risk neutral and there are two signals y_1 and y_2 available for contracting. While one of these signals could be the outcome x itself, this is not necessary. Making the appropriate adjustments to the characterization of the optimal contract, we have

$$\frac{1}{U'[s(y_1, y_2)]} = \lambda + \mu \frac{f_a(y_1, y_2|a)}{f(y_1, y_2|a)}. \tag{9}$$

We can solve for the optimal contract to get

$$s(y_1, y_2) = W\left[\lambda + \mu \frac{f_a(y_1, y_2|a)}{f(y_1, y_2|a)}\right],$$

where W is the inverse of the agent's marginal utility function.

Banker and Datar (1989) were the first to suggest we could decompose the contract into (i) an aggregate performance measure: $\pi = \pi(y_1, y_2) = \lambda + \mu f_a(y_1, y_2|a)/f(y_1, y_2|a)$, and (ii) a compensation function based on the aggregate performance measure, $s(\pi)$.

Using this decomposition, the form of the aggregation process, $\pi(y_1, y_2)$, is determined by the shape of $\lambda + \mu f_a(y_1, y_2|a)/f(y_1, y_2|a)$. Since $\mu > 0$, this is determined by the shape of $f_a(y_1, y_2|a)/f(y_1, y_2|a)$. Banker and Datar show that for many common classes of probability distributions, $f_a(y_1, y_2|a)/f(y_1, y_2|a)$ is

linear in y_1 and y_2.[23] For example, this holds for the exponential family of distributions, which includes the normal, exponential, binomial, gamma, and chi-square. It is important to emphasize that this result only implies that the performance measures are aggregated in a linear fashion; it does not imply that the contract is linear in the performance measures. For example, if the contract is $s(y_1, y_2) = [y_1 + 3y_2]^2$, the signals are aggregated in a linear fashion (e.g., $y_1 + 3y_2$), but the aggregate performance measure is used in a nonlinear way to determine the agent's compensation.

Next we turn to the issue of how much weight do we put on each performance measure. Note that this is well defined when we aggregate the signals in a linear fashion. In particular, the slope coefficient assigned to each variable can be thought of as the weight the variable receives in the contract. For concreteness, let the linear aggregation process be written as $\pi(y_1, y_2) = \beta_1 y_1 + \beta_2 y_2$. Banker and Datar show that for the exponential family of distributions discussed above, the slope coefficients are proportional to the "signal-to-noise" ratio of the variable. By calculating the ratio of the slope coefficients, the proportionality factor cancels out.

Specifically, when y_1 and y_2 are independently distributed given the agent's effort, the ratio of the slope coefficients, or the relative weights in the optimal performance aggregation, satisfies

$$\frac{\beta_1}{\beta_2} = \frac{\partial E(y_1|a)/\partial a}{\text{var}(y_1)} \frac{\text{var}(y_2)}{\partial E(y_2|a)/\partial a}. \tag{10}$$

Note that $E(y_i|a)/\partial a$ is the sensitivity of signal i to the agent's effort. This measures how much the expected value of the signal moves in response to a change in the agent's effort. Eq. (10) indicates that, ceteris paribus, the more sensitive a signal is, the higher the relative weight it receives. Eq. (10) also indicates that the relative weight is decreasing in the variance of the signal. The variance of the signal measures how noisy it is because the variance is driven solely by the importance of other factors (other than the agent's effort) on the signal y. A noisier signal receives a smaller weight, ceteris paribus. Therefore, a signal which is not very sensitive and which is very noisy has a positive weight only because we can adjust the weight to be very small. In contrast, if the contract shape or the magnitude of the weight is exogenously specified, the use of such a signal can lead to a decrease in welfare for the principal and the agent.

Banker and Datar's results provide strong theoretical support for the linear aggregation property of accounting signals. However, their result also suggests

---

[23] Recall that the term f_a/f is the derivative of the log of the likelihood function with respect to a. Therefore, f_a/f is linear in y_1 and y_2 whenever $\partial \log[f(y_1, y_2|a)]/\partial a$ linear in y_1 and y_2. Working it out in reverse, it is linear if $f(y_1, y_2|a) = \exp\{\int g[l(a)y_1 + m(a)y_2]\, \mathrm{d}a + t(y_1, y_2)\}$, because this implies $f_a(y_1, y_2|a)/f(y_1, y_2|a) = g[l(a)y_1 + m(a)y_2]$. Here we implicitly move the g function into the compensation function, as opposed to the performance function.

that the equal weighting principle (i.e., total cost is the sum of individual costs, net income is revenue minus expenses) would rarely be expected to be optimal as a performance measure. That is, any component of revenue or expense that the agent has greater "influence" over should receive higher weight, and any component that was more "volatile" should receive less weight. Only when the signal-to-noise ratios of the components are identical should we expect the "equal weighting" property to be optimal.

However, in richer models, we might expect the weight placed on components of financial performance to be closer than the single-action models suggest. In particular, if the agent has the opportunity to select actions that transfer costs between components, or that increase a revenue and a cost by the same amount, then he will engage in these nonvalue added activities if the components are weighted differently in his compensation function. He may even engage in value-destroying activities to take advantage of different slope coefficients on some components of revenues or costs than others. When these "arbitrage" opportunities are large, the principal is likely to respond to them by equalizing the slope coefficients to reduce the agent's incentives to engage in them.

2.6. The controllability principle and relative performance evaluation

When the signals are independently distributed, any signal which is sensitive to the agent's action is useful in the contract. This seems similar in spirit to the "controllability" principle in accounting. Of course, in a world of uncertainty, the agent does not literally control any performance measures, so the term controllability is a bit of a misnomer. Unfortunately, the conventional management accounting literature does not provide a precise definition of controllability. For purposes of our discussion, we will define a variable as "controllable" if the agent's actions influence the probability distribution of that variable.[24] Using this operational definition, agency theory provides support for the controllability concept if all signals are independently distributed; all variables which are controllable are valuable to include in the agent's compensation.

However, agency theory also shows that it is valuable to include variables in the contract which are not controlled by the agent. While we have not modeled the actions of other agents or the principal yet, it is easy to see how the contract would be affected by these. For example, if the principal makes a capital decision and this affects the mean of the output, the variance of the output, or the productivity of the agent's effort, then the capital decision will show up as a parameter of the contract. For example, if the principal's capital decision

---

[24] See Baiman and Noel (1982), Merchant (1987), and Antle and Demski (1988) for additional discussion of the controllability principle.

affects the mean of a performance measure the principal will subtract this out. If it increases the sensitivity of the agent's effort, the principal will place more relative weight on this signal, etc. Clearly the agent is not being held "responsible" for the principal's capital decision though. Instead the principal is using the capital decision to "adjust" the other performance measures so he can better evaluate the thing the agent does "control"—his actions.

Things are more complicated in the more realistic scenario where the performance measures are correlated. First, variables which are controllable might not be valuable to include in the contract if they are not incrementally informative about the agent's actions. As illustrated earlier, suppose two signals can be written as $y_1 = a + \tilde{e}_1$ and $y_2 = y_1 + \tilde{e}_2 = a + \tilde{e}_1 + \tilde{e}_2$, where $\tilde{e}_1$ and $\tilde{e}_2$ are random variables which are uncorrelated. Clearly, performance measure y_2 is controllable by the agent, but adds no value to the contract because it is merely a garbling of the first performance measure.

Second, there are additional reasons why performance measures the agent cannot influence can be useful in contracting. To see this, when the signals y_1 and y_2 are correlated, Banker and Datar show that the relative weights can be written as

$$\frac{\beta_1}{\beta_2} = \frac{\left[\dfrac{\partial \mathrm{E}(y_1|a)}{\partial a} - \dfrac{\mathrm{cov}(y_1, y_2)}{\mathrm{var}(y_2)}\dfrac{\partial \mathrm{E}(y_2|a)}{\partial a}\right]}{\mathrm{var}(y_1)} \frac{\mathrm{var}(y_2)}{\left[\dfrac{\partial \mathrm{E}(y_2|a)}{\partial a} - \dfrac{\mathrm{cov}(y_1, y_2)}{\mathrm{var}(y_1)}\dfrac{\partial \mathrm{E}(y_1|a)}{\partial a}\right]}. \tag{11}$$

Note that even if y_2 is not sensitive to a, it still might be used in the contract. That is, if $\partial \mathrm{E}(y_2|a)/\partial a = 0$, we have

$$\frac{\beta_1}{\beta_2} = \frac{\left[\dfrac{\partial \mathrm{E}(y_1|a)}{\partial a}\right]}{\mathrm{var}(y_1)} \frac{\mathrm{var}(y_2)}{\left[-\dfrac{\mathrm{cov}(y_1, y_2)}{\mathrm{var}(y_1)}\dfrac{\partial \mathrm{E}(y_1|a)}{\partial a}\right]} = \frac{-\mathrm{var}(y_2)}{\mathrm{cov}(y_1, y_2)}. \tag{12}$$

Eq. (12) shows that both variables receive nonzero weight in the contract as long as their correlation is nonzero. If the two variables are positively correlated, the weight assigned to y_2 has the opposite sign of the weight assigned to y_1. A positive correlation between y_1 and y_2 means they are affected in the same direction by a common exogenous "shock" term. Since the agent's action is not a random variable, this correlation must arise from correlation between their "noise" components. By including y_2 in the contract with a negative weight, some of the noise in the performance measure y_1 can be removed.

This "noise reduction" role of a performance measure has an interesting link to relative performance evaluation and the "single factor" index model

discussed in capital markets research. To see this, let y_1 denote the agent's own performance, and y_2 denote the performance of a peer group. Eqs. (11) and (12) imply that the aggregate performance can be expressed as

$$\beta_1 y_1 + \beta_2 y_2 = \beta_1 \left[y_1 - \frac{\text{cov}(y_1, y_2)}{\text{var}(y_2)} y_2 \right]. \tag{13}$$

The performances of the agent and the peers will often be affected by common random factors because they work in the same environment, are evaluated by the same supervisors, use the same resources or production technology, experience the same macro-economic effects, etc. These common shocks are represented by the covariance between the variables. In particular, suppose we write the agent's performance using a single-factor index model:

$$y_1 = \phi_0 + \phi_1 y_2 + \omega,$$

where the slope coefficient is $\phi_1 = \text{cov}(y_1, y_2)/\text{var}(y_2)$. Using this result, Eq. (13) implies that optimal aggregation of the performance measure is then

$$\beta_1 y_1 + \beta_2 y_2 = \beta_1 [y_1 - \phi_1 y_2] = \beta_1 [\phi_0 + \omega].$$

That is, the "market component" of the agent's performance is removed, and the agent is evaluated solely on the basis of the "idiosyncratic" component of his performance. Note that this performance measure is still very much in the spirit of "controllability". We are using whatever information we can to filter out the effect of all other variables in order to focus on the actions the agent is responsible for. The "informativeness" principle makes it clear that we can do this more effectively by bringing in variables that the agent does not influence.

While relative performance evaluation (RPE) is used in many contexts (e.g., grading on the curve, employee of the month, sports tournaments, etc.), there is very little evidence of its use in executive compensation. A number of potential costs of using relative performance evaluation have been advanced in the literature.[25] For example, there may be counter-productive arguments over what components of performance are "controllable" and what components are noncontrollable. A second problem is that there appear to be "political" costs with shareholder groups when executives are paid large bonuses if their firm's stock price has gone down, even if the decrease is not as large as the decrease for peer firms.

A third cost to evaluating agents relative to a peer group is that it can motivate destructive competition between agents, i.e., making yourself look good compared to a peer group by sabotaging their performances instead of improving your own. This is especially of concern if all the agents are within the same firm, as opposed to comparing the performance of one firm against another. It is not clear why this would be more of a problem at the executive

---

[25] See Dye (1984, 1992) and Fershtman and Judd (1987) for additional discussion.

level (especially at the CEO level) than at lower levels. Fourth, the use of RPE might lead to poor strategic decisions (e.g., picking lines of business where the competition is "easy" as opposed to picking the ones where you will do best on an absolute basis). Including the performance of competitor firms in a compensation package can also affect the type of strategy executives choose to pursue (increase market share versus lower product cost). Fifth, removing the impact of a variable from the agent's performance measure reduces his incentives to forecast that variable and modify the firm's strategy on the basis of this information. For example, even if oil prices are exogenous to given executive, we may still want the executive to attempt to forecast what oil prices will be and to design a strategy for the firm that is best given that strategy (inventory decisions, pricing contracts, hedging positions, etc.). Finally, it is possible that executives can achieve some the benefits of RPE on their own. In particular, they may be able to re-allocate their portfolio of wealth to remove a portion of the market-related risk. As a result, it is unnecessary for the firm to do with the compensation contract. To explore this last possibility, it is necessary to explicitly model the agent's outside portfolio of wealth and his investment opportunity set.

2.7. Magnitude of the value of a performance measure

Agency theory has derived conditions where information has nonzero value and has examined the factors that affect the weight assigned to performance measures. However, relatively little attention has been directed at how *much* value the performance measure has. One problem with interpreting the weight assigned to a performance variable as a measure of its value is that the weight is obviously affected by the scale of the variable. That is, re-scaling the variable by multiplying it by two will cause the weight on the variable to be cut in half, though nothing of substance will have changed. One possible solution is to re-scale the variables, say to have the same sensitivity or the same variance.[26] Even if the variable is re-scaled, it is not clear whether there is a link between the weight assigned to the performance measure and the value added by the performance measure. There has been surprisingly little attention devoted to this issue.

An exception is Kim and Suh (1991).[27] They examine the case where the agent has a square root utility function, $U(s) = \sqrt{s}$. This utility function has proven to be the most tractable in solving agency theory problems. Consider

---

[26] See Lambert and Larcker (1987), Sloan (1993), Lambert (1993), and Feltham and Wu (2000) for additional discussions of the scaling of the performance measures.

[27] See also Baiman and Rajan (1994), who examine the value of information systems as a function of their Type I and II errors. Rajan and Sarath (1997) examine the value of multiple information signals when the signals are correlated.

two competing information systems: one generates a signal y_1 which has density function $f^1(y_1|a)$ and the other generates a signal y_2 which has density function $f^2(y_2|a)$. Kim and Suh (1991) show if the principal wants to motivate a given action a, then information system one is preferred to information system two, if and only if

$$\text{var}\left[\frac{f_a^1(y_1|a)}{f^1(y_1|a)}\right] > \text{var}\left[\frac{f_a^2(y_2|a)}{f^2(y_2|a)}\right].$$

In the statistics literature, the variance of $f_a^i(y_i|a)/f^i(y_i|a)$ is referred to as the "amount of information" that the signal conveys about the action.[28] Note that a higher variance of $f_a^i(y_i|a)/f^i(y_i|a)$ is not the same thing as a higher variance for y_i. In fact, for the exponential family of distributions, recall that we have

$$\frac{f_a^i(y_i|a)}{f^i(y_i|a)} = \frac{\partial E(y_i|a)/\partial a}{\text{var}(y_i)}(y_i - E(y_i|a)).$$

This implies

$$E\left[\frac{f_a^i(y_i|a)}{f^i(y_i|a)}\right]^2 = \frac{(\partial E(y_i|a)/\partial a)^2}{\text{var}(y_i)}.$$

This is similar to the sensitivity-to-noise ratio in Banker and Datar. The only difference is that the sensitivity squared appears in the numerator. This makes the measure of the value of the information system independent of the scale of the performance measure. Other things equal, an information signal has more value if it has a higher sensitivity and a lower variance.

Surprisingly little work has been done to establish the properties of the magnitude of the value of single performance measures in more complicated models (multiple actions or private information) or in establishing the magnitude of the *incremental* value a signal has when there are other information signals already available for contracting.[29]

3. Multi-action models

While single-action agency models have been useful in generating many insights, they are too simple to allow us to address some important features of performance measures. In particular, in single-action models the sensitivity of a signal is an important feature, but the single-action framework precludes us from asking whether the measure is sensitive to the "right things". In reality, we know that agents are generally responsible for a rich set of actions. They

---

[28] A similar result is found by Lambert (1985) in a variance investigation setting.

[29] See Feltham and Xie (1994) and Banker and Thevaranjan (2000) for analyses of the value of a performance measure in a multi-action setting.

can vary how much attention they spend on one product line versus another, on generating revenues versus decreasing costs, on customer satisfaction or product quality, on design versus operations, on new investment, etc. Moreover, we know that not all performance measures are equally sensitive to a particular action, some can be more easily manipulated than others, some reflect information on a more timely basis, etc. For example, the idea of the balanced scorecard is an attempt to capture the multi-dimensionality of agents' actions and the differential ability of performance measures to reflect these actions and their results.

Conceptually, there are no additional difficulties to extending the Holmstrom framework to multi-action settings. Assume we can represent the agent's choice of efforts using his first-order conditions and let μ_i be the Lagrange multiplier on the first-order condition associated with action $i(i = 1, \ldots, m)$. Analogous to Eq. (9), the optimal contract can be characterized as

$$\frac{1}{U'[s(\mathbf{y})]} = \lambda + \mu_1 \frac{f_{a_1}(\mathbf{y}|\mathbf{a})}{f(\mathbf{y}|\mathbf{a})} + \mu_2 \frac{f_{a_2}(\mathbf{y}|\mathbf{a})}{f(\mathbf{y}|\mathbf{a})} + \cdots + \mu_m \frac{f_{a_m}(\mathbf{y}|\mathbf{a})}{f(\mathbf{y}|\mathbf{a})}. \tag{14}$$

Banker and Datar's (1989) results on when linear aggregation is optimal continue to apply, and the relative weights on the performance measures in the contract continue to have a "sensitivity–precision" interpretation. While the definition of the precision of a signal is the same as in one-action model (e.g., the variance), the sensitivity is more complicated. In particular, in a multi-action model, the overall sensitivity of a performance measure is a weighted sum of its sensitivities to the individual actions. The weights applied to these sensitivities are the Lagrange multipliers on the incentive compatibility constraints, μ_i. Since the Lagrange multiplier represents the marginal impact on the principal's net profits of relaxing the constraint, this result implies that the overall performance measure weights each individual sensitivity by its "importance" in the incentive problem.

In the single-action result derived in Banker and Datar (1989), there is only a single Lagrange multiplier on the agent's effort, so this Lagrange multiplier cancels in the calculation of the relative weights assigned to the performance measures. As a result, no other features of the model affect their relative weights except the sensitivity and the variance of the performance measures. As Eq. (14) suggests, if the μ_i are unequal, they do not cancel in the calculation of the ratio of the weights in our multi-action model.[30] This opens the possibility for other features of the model to affect the relative weights. Unfortunately, the Lagrange multipliers, which are endogenous variables to the model, are difficult to solve for, which makes examination of these issues problematic. In

---

[30] Similarly, Bushman and Indjejikian (1993a) find that the Lagrange multipliers do not cancel in their multi-action model.

the following sections, we place additional structure on the model to enable us to further characterize the factors that affect the relative weights assigned to the performance measures.

3.1. Linear–exponential–normal (LEN) formulation of agency models

An alternative formulation of the agency model, developed by Holmstrom and Milgrom (1987) has proved to be much more tractable in addressing multi-action and multi-period models. However, this tractability is achieved by severely restricting the generality of the model along three dimensions. First, the agent's utility function is assumed to be negative exponential, $U(w) = -\mathrm{e}^{-\rho w}$, where ρ is the agent's coefficient of risk aversion. The important feature of this utility function is that it exhibits constant absolute risk aversion. This means that the agent's wealth does not affect his risk aversion and therefore does not affect the agent's incentives. This is especially important in multi-period models, where the agent's wealth will generally vary over time. Second, the performance measures are assumed to be normally distributed. The primary advantage of normality is that the mean can be affected without affecting higher moments of the distribution. Finally, the compensation functions are assumed to be linear in the performance measures. The combination of these three assumptions means that the agent's expected utility (more accurately, the certainty equivalent of the expected utility) has an especially tractable form.

The restriction to linear contracts has been a controversial one in the agency literature. In particular, agency theory's intellectual roots lie in information economics, where information systems are compared based on the *optimal* use of information generated by the system. Restricting the contract to be linear is a very significant philosophical departure because, as the earlier sections have demonstrated, in single-period models linear contracts are rarely optimal (though linear aggregation of performance measures frequently is).

There are three common justifications for the linear contracting restriction. One motivation is that these linear contracts are, in fact, optimal in a richer (but typically unmodeled) setting. Holmstrom and Milgrom (1987) (HM) develop a continuous time model in which the agent affects the drift rate of a Brownian motion process. Even though the agent is allowed to dynamically adjust his effort over time, HM show that the optimal solution is equivalent to one in which the agent selects a single-effort level and the principal restricts himself to a linear contract. By "pretending" our one-period models are "snapshots" within a continuous time model such as that modeled by HM, the linear contracting framework has strong theoretical justification. Unfortunately, it is not clear that the HM framework extends to models with multiple actions and multiple performance measures, especially if the performance

measures are correlated or if there are more actions than performance measures.

A second motivation for using the linear contracting framework is that linear contracts are commonly used in practice. I find this explanation less convincing. Of course, there are some settings where a linear contract is a good approximation to the *explicit* contract used. However, once the implicit incentives, the judgment and discretion in compensation decisions, and the internal and external "appeals" process are taken into consideration, I suspect linearity is less descriptive of the "total" contract. Moreover, even if we confine our attention to the explicit contract, I believe these contracts are more likely to be piecewise linear than linear.

The third motivation for using linear contracts is simply their tractability. I believe this is the real reason researchers have moved to linear contracts. "Conventional" agency theory models using the Holmstrom (1979) framework cannot be pushed very far in any of the directions accounting and economics researchers are interested in exploring before we reach the limits (or certainly the point of extremely diminishing marginal returns) in our ability to solve for the optimal contracts. Researchers realized that sacrifices would have to be made in order to use this framework to address more realistic and interesting problems. The linear contract restriction allows us to solve much more complicated and interesting models.

Should we be willing to sacrifice "optimal" contracting for tractability? The answer depends on the question the researcher is interested in. There are some research questions that we simply cannot address by assuming linear contracts (e.g., questions about contract shape). Moreover, there are questions where the shape of the contract significantly affects the incentives involved (e.g., earnings management across periods or risk taking). However there are other questions where the assumption of linearity seems less likely to dramatically affect the qualitative conclusions reached; it merely simplifies the math. For example, the results on aggregation of signals into contracts and the congruity of performance measures seem to be robust to the LEN specification. The research question (and available research technology) should lead us to judgments about where to build the detail into the model, and what to regard as exogenous versus endogenous.

3.2. Multi-action models using the LEN framework

The agent is modeled as being responsible for m-dimensions of effort, denoted $\mathbf{a} = (a_1, a_2, \ldots, a_m)$. The principal cannot observe the individual effort levels or the total effort. In later sections we will provide more concrete applications of the analysis to specific types of actions. Let x denote the firm's end-of-period cash flow (before compensation to the agent). The firm's outcome is not necessarily observable to the principal. Instead, the principal

must base the agent's compensation on K signals, $\mathbf{y} = (y_1, y_2, \ldots, y_K)$. Note that the number of actions can exceed the number of performance measure available for contracting. When this is the case, the principal's control problem is more difficult because he has fewer "control" variables (slope coefficients) than variables to control (actions).

The performance measures $\mathbf{y}$ are assumed to be distributed as normal random variables. The agent's actions are assumed to affect the means but not the variances or covariances of the performance measures. For convenience, the expected values of x and $\mathbf{y}$ are all assumed to be linear functions of the agent's efforts. The linearity assumption here is not critical; it just makes the math easier. As discussed above, the assumption of normality is much more important. Let the outcome function be

$$x = \sum_{j=1}^{m} b_j a_j + e_x,$$

and the performance measures be

$$y_i = \sum_{j=1}^{m} q_{ij} a_j + e_i \quad \text{for } i = 1, \ldots, K.$$

The sensitivities of the performance measures to actions, q_{ij}, need not equal the sensitivity of the outcome x to those actions.

For example, one performance measure could be the firm's "true" outcome, x, in which case the vector of q's would be identical to the vector of b's. Alternatively, different performance measures could be components of the firm's outcome: components of costs or revenues, or divisional profits. Different actions could then affect some components of performance but not others. In the divisional income interpretation, the agent could be thought of as allocating his effort between activities that improve the profitability of his own division and activities that improve (or make worse) the profitability of other segments of the firm.[31] A measure like accounting earnings could be represented as fully capturing the financial effect of "short term" operating decisions but not completely reflecting the long-term profitability of actions such as product development, quality, or customer service. A nonfinancial measure might be the one that is affected by these actions but which also contains considerable measurement error or noise.[32]

---

[31] See Bushman et al. (1995) for analysis of this issue in a single-action setting where the agent's effort has spillover effects to other divisions and Datar et al. (2000) for analysis of these issues in a multi-action setting.

[32] See Thevaranjan et al. (2000) and Hemmer (1996) for analyses of quality and customer satisfaction interpretations of the model. Hemmer (1996) models the performance measures as multiplicative in the agent's actions and the disturbance terms as log-normally distributed. However, he conducts his analysis using the logs of the performance measures, which converts his model to the additive, linear, normal distribution assumptions of our model.

The variables e_x and e_i are jointly normally distributed with zero means. As in the single-action models, these variables represent the impact of all factors other than the agent's actions on the outcome and the performance measures. We let $\text{var}(y_i)$ denote the variance of y_i and we let $\text{cov}(y_i, y_j)$ denote the covariance between performance measures i and j. For completeness, let $\text{cov}(x, y_i)$ be the covariance between the outcome and performance measure i. It will turn out that the covariance between performance measure i and the outcome x is not important from a contracting perspective unless x is one of the performance measures available for contracting. However, as we discuss in a later section, such a correlation is extremely important from a valuation perspective.

We assume the agent's utility function displays constant absolute risk aversion, which eliminates wealth effects on the optimal contracting problem. Specifically, the agent's utility function is of the negative exponential form, $U(W) = -e^{-\rho W}$, where ρ is the coefficient of absolute risk aversion and $W = s(\mathbf{y}) - V(\mathbf{a})$ is the agent's "net" income after deducting the "cost" of effort. That is, V represents the monetary equivalent of the agent's disutility of effort. Consistent with much of the agency literature that has followed Feltham and Xie (1994), we assume the agent's cost of effort is additively separable into its components, $V(\mathbf{a}) = 0.5 \sum_{j=1}^{m} a_j^2$. Note that the agent is both effort and (weakly) risk averse. In contrast, the principal is assumed to be risk neutral.

The compensation contract is assumed to be a linear function of the observed performance measures:

$$s(y_1, \ldots, y_K) = \beta_0 + \sum_{i=1}^{K} \beta_i \tilde{y}_i.$$

The combination of a normal distribution, linear contract, and negative exponential utility implies that the certainty equivalent (CE) of the agent's expected utility can be expressed in a convenient form as follows:[33]

$$\begin{aligned}
\text{CE} &= \text{E}[s(\mathbf{y})] - 0.5\rho \, \text{var}[s(\mathbf{y})] - V(\mathbf{a}) \\
&= \beta_0 + \sum_{i=1}^{K} \beta_i \text{E}(y_i | \mathbf{a}) - 0.5\rho \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j \, \text{cov}(y_i, y_j) \right] - V(\mathbf{a}).
\end{aligned}$$

Note that the agent's certainty equivalent is simply his expected compensation minus a cost for bearing risk, which depends on the variance of his compensation and how risk averse he is, minus the cost of his effort.

---

[33] The certainty equivalent is the sure payment (net of the cost of effort), CE, that provides him with the same expected utility as he will receive with the risky contract, $s(\mathbf{y})$. Therefore, CE satisfies $U(\text{CE}) = \text{E}[U[s(\mathbf{y}) - V(\mathbf{a})]]$.

With this structure, the principal's problem can be expressed as

$$\max_{\beta_0,\beta_1,\ldots,\beta_K,\mathbf{a}} \quad E(x|\mathbf{a}) - E[s(\mathbf{y})]$$

subject to   (AUC) $E[s(\mathbf{y})] - 0.5\rho \, \text{var}[s(\mathbf{y})] - V(\mathbf{a}) \geqslant CE(\underline{H}),$

(AIC) $\mathbf{a}$ maximizes $E[s(\mathbf{y})] - 0.5\rho \, \text{var}[s(\mathbf{y})] - V(\mathbf{a}).$

The principal maximizes his expected profits net of the agent's compensation. The agent's acceptable utility constraint (AUC) requires the contract to be sufficiently attractive to induce the agent to accept it. The agent's expected utility is expressed using its certainty equivalent. The certainty equivalent of the agent's reservation utility, $CE(\underline{H})$ is, without loss of generality, assumed to be zero. The agent's incentive compatibility constraint (AIC) requires the actions to be in the agent's best interests given the compensation contract offered. The incentive contract must motivate the desired allocation of effort, as well as the desired total level of effort.

The intercept of the compensation contract, β_0, can be chosen to make the (AUC) constraint an equality.[34] We can therefore eliminate this constraint by substituting this expression directly into the objective function, as follows:

$$\max_{\beta_1,\ldots,\beta_K,\mathbf{a}} \quad \sum_{j=1}^{m} b_j a_j - V(\mathbf{a}) - 0.5\rho \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j \, \text{cov}(y_i, y_j) \right]$$

subject to (AIC) $\mathbf{a}$ maximizes $\beta_0 + \sum_{i=1}^{K} \beta_i E(y_i|\mathbf{a}) - V(\mathbf{a}).$

Note that since the agent's actions do not affect the variances or covariances of the performance variables, the variance term drops out of the (AIC) constraint.

We can think of the contract as weighting the performance measures so that the overall performance measure used to evaluate the agent is $\beta_0 + \sum_{i=1}^{K} \beta_i E(y_i|\mathbf{a})$. This can be expanded as

$$\beta_0 + \sum_{i=1}^{K} \beta_i \left[\sum_{j=1}^{m} q_{ij} a_j \right] - 0.5 \left[\sum_{j=1}^{m} a_j^2 \right]$$

$$= \beta_0 + \sum_{j=1}^{m} \left[\sum_{i=1}^{K} \beta_i q_{ij} \right] a_j - 0.5 \left[\sum_{j=1}^{m} a_j^2 \right].$$

Note that the sensitivity of the agent's expected compensation to action j is $\sum_{i=1}^{K} \beta_i q_{ij}$, compared to the sensitivity of the expected outcome to that action being b_j.

---

[34] The intercept is chosen to make $E[s(\mathbf{y})] = 0.5\rho \, \text{var}[s(\mathbf{y})] + V(\mathbf{a}) + CE(\underline{H})$. Note with a linear contract and normal distribution, we are allowing the agent's payments to be negative, and unboundedly so.

With this structure the agent's first-order conditions on his actions simplify to[35]

$$\sum_{i=1}^{K} \beta_i q_{ij} - a_j = 0, \quad j = 1, \ldots, m. \tag{15}$$

Substituting these equations for the agent's actions into the principal's objective function allows us to formulate the principal's problem without constraints:

$$\max_{\beta_1, \ldots, \beta_K} \sum_{j=1}^{m} b_j \left[\sum_{i=1}^{K} \beta_i q_{ij} \right] - 0.5 \sum_{j=1}^{m} \left[\sum_{i=1}^{K} \beta_i q_{ij} \right]^2 \\ - 0.5\rho \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j \operatorname{cov}(y_i, y_j) \right]. \tag{16}$$

This optimization problem is tractable enough that it can be solved in closed form for the optimal weights in the compensation contract. However, before doing this, Datar et al. (2000) show that it is useful to rewrite the objective function as follows:

$$\underset{\beta_1, \ldots, \beta_K}{\text{minimize}} \sum_{j=1}^{m} \left[b_j - \sum_{i=1}^{K} \beta_i q_{ij} \right]^2 + 0.5\rho \left[\sum_{i=1}^{K} \sum_{j=1}^{K} \beta_i \beta_j \operatorname{cov}(y_i, y_j) \right]. \tag{17}$$

As we discuss below, the first term reflects the principal's desire to motivate the right allocation of the agent's efforts. In particular, these terms cause the performance measures to be weighted to maximize the *congruence* between the firm's outcome and the agent's overall performance measure. The second term reflects the principal's desire to reduce the risk in the agent's compensation, because the principal must increase the agent's expected compensation to offset the disutility associated with his bearing risk.

To isolate the effect of the first term, assume the agent is risk neutral (i.e., $\rho = 0$) or the variances of the performance measures equal zero. When the agent is risk neutral or the performance measures are noiseless, risk considerations are unimportant and the last term in Eq. (17) vanishes. Nevertheless, there is still a nontrivial performance measurement problem because the outcome x is unobservable. That is, the noncontractability of the outcome x precludes the principal from simply selling the firm to the agent. The principal's problem is to design the contract to make the agent's overall performance measure $\sum_{i=1}^{K} \beta_i y_i$ as congruent as possible to the firm's outcome

[35] A sufficient condition to justify the use of the first-order condition approach is that the agent's expected utility be a strictly concave function of his actions. Since the performance measures are linear functions of the agent's efforts and the cost of effort is convex, this makes the agent's expected utility strictly concave.

(x). Eq. (17) indicates that the incongruity of the agent's overall performance to the firm's outcome is

$$\sum_{j=1}^{m} \left[b_j - \sum_{i=1}^{K} \beta_i q_{ij} \right]^2 . \tag{18}$$

Under limited conditions, perfect congruity can be achieved. Specifically, the matrix of the performance measure sensitivities q_{ij} has to be such that the sensitivities can be combined to exactly replicate the outcome sensitivities to all of the agent's actions, b_j. That is, there must exist a $(\beta_1, ..., \beta_K)$ such that $\sum_{i=1}^{K} \beta_i q_{ij} = b_j$ for all $j = 1, ...m$. For example, this can always be achieved in a single-action model by simply rescaling the performance measure to make $\beta_1 q_{11} = b_1$. More generally, if there are at least as many performance measures as there are actions, perfect congruity can be achieved as long as the vectors of performance measure sensitivities are nondegenerate and not linearly dependent. However if there are more actions than performance measures, perfect congruity cannot be achieved except by "accident" because the principal has fewer control variables (i.e., slope coefficients) than he has variables to control (i.e., actions).

When perfect congruity cannot be achieved, the principal weights the measures to make them "as congruent" as possible. Eq. (18) shows that the degree of congruity is measured by the (square of the) distance between the vector of the outcome's sensitivity to the agent's actions (b_j) and the vector of the sensitivity of the agent's overall performance measure to these actions ($[\sum_{i=1}^{K} \boldsymbol{\beta}_i q_{ij}]$). This distance is given by $\sum_{j=1}^{m} [b_j - \sum_{i=1}^{K} \beta_i q_{ij}]^2$.

To achieve maximal congruity, we can think of the principal as first choosing the weights in the compensation contract to minimize the angle, θ, between the vector of outcome sensitivities b_j and the vector of the agent's overall performance measure, $[\sum_{i=1}^{K} \beta_i q_{ij}]$. The principal then scales this overall vector to make it as close as possible to $\mathbf{b}$ by specifying the strength of the incentive placed on the overall performance measure.[36]

The second effect on the optimal weights in Eq. (17) is a sensitivity–precision effect. To focus on this effect, assume the performance measures are perfectly aligned *with each other* (i.e., there exists a "base" vector $\mathbf{q}_j$ and constants $\phi_i \neq 0$ such that $q_{ij} = \phi_i q_j$ for $j = 1, ..., m$). Under these conditions, the angle (θ) between the agent's overall performance measure and the firm's outcome is pre-determined, and the only choice variable is the length. Therefore, the weights assigned to the performance measures cannot affect the allocation of

---

[36] In fact, the weights in the compensation contract are the same as would be obtained from regressing (with no intercept) the coefficients b_j on the coefficients $(q_{1j}, ..., q_{Kj})$. As in multiple regression, the weight assigned to a performance measure is not simply a function of its own "congruence" with the outcome, but also on how it interacts with the other variables in the contract.

the agent's effort; the compensation contract can only be used to affect the overall *intensity* of the agent's effort.[37]

In this case, the model is substantively the same as a single-action model. Therefore, the optimal weights are chosen based on the (covariance adjusted) sensitivity and noise of the performance measures, exactly as in Banker and Datar's single-action model. In this case, the overall sensitivity of the signal y_i reduces to the parameter ϕ_i.

More generally, both the congruity effect and the sensitivity–precision effect are present. For the case of two performance measures, we can solve Eq. (17) for the ratio of the optimal weights to be

$$\frac{\beta_2}{\beta_1} = \frac{\left[\sum_{j=1}^m b_j q_{2j}\right]\left[\sum_{j=1}^m q_{1j}^2\right] - \left[\sum_{j=1}^m b_j q_{1j}\right]\left[\sum_{j=1}^m q_{1j}q_{2j}\right] + \rho\,\mathrm{var}_1\sum_{j=1}^m f_j g_{2j} - \rho\,\mathrm{cov}\sum_{j=1}^m b q_{1j}}{\left[\sum_{j=1}^m b_j q_{1j}\right]\left[\sum_{j=1}^m q_{2j}^2\right] - \left[\sum_{j=1}^m b_j q_{2j}\right]\left[\sum_{j=1}^m q_{1j}q_{2j}\right] + \rho\,\mathrm{var}_2\sum_{j=1}^m f_j g_{1j} - \rho\,\mathrm{cov}\sum_{j=1}^m b_j q_{2j}}.$$

$$(19)$$

The first two terms in the numerator and denominator are the congruity effect (because when $\rho = 0$ or the performance measures have zero variance, the last two terms in the numerator and denominator vanish). Similarly, the last two terms in the numerator and denominator of Eq. (19) are the sensitivity–precision effect (because the first two terms in the numerator and denominator vanish when the two measures are perfectly aligned with each other).

3.3. Applications of multi-action results

Next we turn to a number of applications of the results of multi-action models. We begin with the situation where there is a single performance measure, then look at multiple performance measures. To be concrete, suppose we have a two-action model, but the principal has only a single performance measure, y_1, available for contracting. In this case, the slope coefficient on the performance measure can be solved to be

$$\beta_1 = \frac{b_1 q_{11} + b_2 q_{12}}{q_{11}^2 + q_{12}^2 + \rho\,\mathrm{var}(y_1)}.$$

$$(20)$$

In the next two sections, we discuss common ways that the performance measure can deviate from measuring the real outcome and the impact this has on the use of the performance measure in the slope control.[38]

----

[37] Note that this result does not depend on whether the two performance measures are congruent with the firm's outcome, only that they are congruent with each other.

[38] The analysis in these two sections builds on the analysis in Feltham and Xie (1994).

3.3.1. Window dressing and when incentives are harmful

One common way in which performance measures can deviate from the real outcome is that the agent can take actions to increase the reported measure without increasing the real outcome. This activity can be thought of as "performance padding", "window dressing", or even lobbying supervisors (Milgrom and Roberts (1992), refer to these as influence costs). It can also be considered a form of earnings management. Let action 1 be "real" (i.e., b_1 and q_{11} are both positive), but action 2 be just window dressing ($b_2 = 0$, but $q_{12} > 0$).

In this case, Eq. (20) indicates that the more susceptible the performance measure is to window dressing (i.e., the higher is q_{12}), the lower the slope coefficient on the measure is. The principal lowers the slope coefficient to discourage the agent from taking action a_2 because the principal must compensate the agent for his effort devoted toward window dressing activities, even though it is not productive. The principal cannot place zero weight on the performance measure though, because it is the only performance measure available to induce productive action a_1. The principal must therefore trade off the benefits of inducing more productive actions (a_1) with the cost of inducing more window dressing (a_2) in deciding how much to use the performance measure.

Things are even worse if the "window dressing" actions actually have a negative impact on the real outcome. Under this scenario ($q_{12} > 0$, but $b_2 < 0$), the costs of using the performance measure are higher, and the slope coefficient is even lower. It is now possible that the optimal slope coefficient is zero, so that it is better to give the agent no incentives rather than give him "bad" incentives.[39] Clearly there would be benefits to obtaining a second performance measure that could be used to untangle the effect of "production effort" from "performance padding".

It is important to note that the "window dressing" actions modeled here take place at the *beginning* of the period, before the agent has observed the "preliminary" or "real" outcome. However, due to the special structure of the model, the result is exactly the same as if the agent waited till the *end* of the period and observed the outcome before deciding how much effort to devote towards window dressing. That is, because the contracts are (exogenously assumed to be) linear and the agent's risk aversion is independent of his wealth level, his decision regarding how much effort he should devote toward window dressing does not depend on the realization of the performance measures. Therefore, models with this special structure are not well suited for examining or explaining earnings management activities where the agent's decision to

---

[39] While it is also possible that the solution could be to make the slope coefficient negative, this seems unlikely if the amount of window dressing is constrained to be nonnegative.

manage earnings up or down depends on the realization of the performance measure.[40]

3.3.2. Myopic performance measures and adding a performance measure

A performance measure can also be incongruent because it is not sensitive to all of the real effects of the agent's actions. For example, accounting net income or operating cash flow measures generally do not immediately reflect the long-term impact of actions like capital investments, research and development, or actions that improve customer satisfaction, quality, etc. Similarly, divisional profits may not reflect the spillover effect of actions on the profits of divisions. We can model extreme performance measure "myopia" by assuming $b_2 > 0$ (i.e., action two has an effect on the real outcome), but $q_{12} = 0$ (it has no effect on the performance measure). In this case, the performance measure cannot be used to motivate the second action at all. The slope coefficient is chosen solely to motivate the first action.[41] The same result would hold in the more extreme case where $b_2 > 0$ but $q_{12} < 0$. That is, the performance measure moves in the opposite direction as real outcome. This situation can arise if the agent's action has long-term benefits, but the available performance measure picks up only the upfront costs and none of the future period benefits of this action. Expensing research and development costs or internally created intangible assets would be examples.

In order to motivate the second action, the principal must either "correct" the first performance measure for its "incompleteness" or to supplement the first performance with a second one that is sensitive to the agent's second action. Transfer pricing is an example of the first solution, where ideally the transfer price causes the agent to consider the impact of his actions on other divisions. Nonfinancial performance measures are example of the second solution.

In choosing additional performance measures it is important to consider the properties of the existing performance measures. It is not enough to know

---

[40] See Demski et al. (1984) and Verrecchia (1986) for comparisons of earnings management and accounting choice ex ante versus ex post.

[41] Holmstrom and Milgrom (1991) develop an example where basing compensation on a myopic measure is actually worse than providing no incentives. This does not happen in the model discussed here because of two key modeling differences. First, HM assume the agent cares about the sum of his efforts but not its allocation. Second the agent is assumed to be willing to provide some amount of effort on all dimensions even if he is given no incentives. For this reason, when the agent is given no explicit incentives he is willing to provide a positive amount of overall effort and to allocate this effort across the actions in whatever way the principal desires. When the principal introduces a myopic performance measure, the agent's overall effort increases, but all of it gets skewed toward the one action to which the myopic measure is sensitive. If a little of all the actions yields more profits than a lot of one action, then the principal is better off offering no incentives at all.

"how" incongruent they are, but also "where" they are incongruent, or "in what direction" they are incongruent. Intuitively, the principal would like to choose new performance measure which are incongruent in the opposite way as the existing ones in order to be able to construct an overall performance measure that is as congruent as possible. To date, this remains a relatively unexplored topic.[42]

3.3.3. Tradeoffs between congruity and sensitivity–precision

In this section, we discuss the tradeoffs between congruity and sensitivity/precision concerns. Specifically, we examine the special case where one performance measure is perfectly congruent with the firm's outcome, while the other measure is not. In this situation, any weight placed on the noncongruent measure means that the principal is willing to sacrifice some congruity in order to achieve better risk sharing. Let the first performance measure, y_1, be identical, and therefore perfectly congruent, to the firm's outcome, $x = y_1 = \sum_{j=1}^{m} b_j a_j + e_1$. In contrast, the second performance measure is an "incomplete" or "local" measure of performance; it is affected by only a subset of the agent's actions. Specifically, let $y_2 = q_{21} a_1 + e_2$. The sensitivity of y_2 to a_1 is q_{21}, while its sensitivity to all other actions is zero.[43]

Substituting this structure into Eq. (19) allows us to express the relative weights in the optimal contract as

$$\frac{\beta_2}{\beta_1} = \frac{b_1 q_{21} \rho \, \text{var}_1 - \rho \, \text{cov} \sum_{j=1}^{m} b_j^2}{q_{21}^2 \sum_{j=2}^{m} b_j^2 + \rho \, \text{var}_2 \sum_{j=1}^{m} b_j^2 - b_1 q_{21} \rho \, \text{cov}}. \tag{21}$$

This equation indicates that the congruent measure (y_1) does not generally receive all the weight in the contract. Since the contract does not place all the weight on the congruent measure, the agent's overall performance measure ($\beta_1 y_1 + \beta_2 y_2$) is not perfectly congruent with x. The principal is willing to sacrifice some of the congruity of the overall performance measure used in the contract in order to reduce the cost of compensating the agent for his effort and for bearing the risk the contract imposes on him. For example, if b_1 and q_{21} are positive, the congruent measure is noisy ($\text{var}_1 > 0$), and the signals are uncorrelated, local measure y_2 receives positive weight in the contract.

As b_j, the productivity of a_j, increases for $j > 2$, the relative weight placed on the congruent signal y_1 increases. When the overall outcome becomes more sensitive to an action, the principal wants to motivate the agent to increase this action. Since the local measure y_2 is not sensitive to these higher dimensions of

---

[42] See Bushman and Indjejikian (1993a, b) for arguments in favor of generating "distorted" performance measures.

[43] The results in this section also hold if the second performance measure is affected by all m actions. However, the tradeoffs between congruence and risk sharing are clearer for the special case examined in the text.

effort, it cannot be used to increase them. Therefore, the contract places more weight on the congruent measure y_1 to motivate the more productive dimension of effort.[44]

However, increasing the other sensitivities (either b_1 or q_{21}) has a nonmonotonic effect on the relative weights in the contract. That is, in contrast to the single-action case, an increase in the sensitivity of a performance measure does not necessarily result in an increase in the relative weight assigned to that performance measure. In the single-action case, the only incentive problem concerns the intensity of the agent's effort. Any variable sensitive to the agent's effort can be used to increase that effort. However, in the multi-action case, the contract must not only motivate the overall intensity of the agent's effort, but also its allocation across its dimensions.

For example, as q_{21} increases, putting more weight on y_2 takes advantage of the signal-to-noise ratio effect of Banker and Datar, but eventually the incongruity of y_2 to the firm's outcome also increases. If the sensitivity of y_2 increases too much, the principal cannot continue to increase the weight in the contract on y_2 or else the agent will provide more of the first dimension of effort than is cost effective. As a result the principal eventually stops increasing the weight assigned to y_2. He uses the increase in the sensitivity of y_2 to motivate the agent to supply a_1, and increases the relative weight on y_1 to motivate the agent to supply the other dimensions of effort.

The nonmonotonic effect of increasing b_1 is more surprising because this sensitivity has a direct productive effect, not just an information effect. As Datar et al. (2000) explain, the reason for the nonmonotonic effect is the principal's concern about incentive spillovers to other actions. That is, if the principal increases the slope coefficient on y_1, this will not only induce the agent to increase a_1 (which is what the principal wants), but will also motivate him to increase the other dimensions of effort (whose productivities have not increased). On the other hand, if the principal uses measure y_2 to provide the increased incentive to supply more a_1, he does not take advantage of the increased sensitivity of signal y_1; the sensitivity of signal y_2 has not increased. However, he avoids the cost of having to compensate the agent for more of the other dimensions of effort. These results suggest that the interaction between incentive effects on different dimensions can be very complicated. Exploring these interactions in other settings would be of great interest.

3.3.4. Divisional versus firm-wide performance

To analyze this application, let performance measure y_1 be the profits for the firm as a whole, and let y_2 be the profits of the agent's division (gross of

---

[44] Since this increase in the weight placed on measure y_1 also increases the agent's incentives to provide more a_1, there is less need to use the local measure y_2 in the contract. Therefore, the weight placed on measure y_2 decreases. The increase in β_1 and the decrease in β_2 both work in the direction of decreasing the relative weight assigned to the local measure y_2.

compensation in both cases). Let the first dimension of effort, a_1, represent effort that improves the profits of the agent's own division and possibly spills over the other divisions. That is, by allowing q_{21} to be different from b_1, the first action, a_1, can have a "spillover" effect on company-wide profits. The expected impact of a_1 on the agent's own division is $q_{21}a_1$ and the expected impact on other divisions is $(b_1 - q_{21})a_1$, so its total impact is b_1a_1. Note that if $b_1 > q_{21}$, then a_1 has a positive spillover (as in Bushman et al., 1995). If $b_1 < q_{21}$, then a_1 has a negative spillover, i.e., it helps his own division more than it helps the firm. In contrast, the second dimension of effort, a_2, represents effort that has no impact on the agent's own divisional profits, but does affect company-wide results (e.g., the profits of other divisions).[45] The incentive problem is therefore one of motivating the agent to select an optimal allocation of effort between "local" and "external" profit-enhancing activities.

Intuitively, one might expect that as the amount of "interdependencies" between the divisions increases, the weight applied to firm-wide profits versus divisional profits would increase. In fact, in a single-action setting this can be shown to be true (see Bushman et al., 1995). However, in a multi-action model, the result depends on the type of interdependency. In particular, as b_1 increases (so the spillover effect of action 1 increases), the relative weight placed on the firm-wide performance measure first decreases, then increases. However, as the second type of interdependency between the divisions increases (i.e., as b_2 increases), there is an unambiguous increase in the relative weight assigned to the firm-wide measure. Again, exploring the interaction between incentive effects seems a fruitful area for future research.

3.3.5. Stewardship versus valuation uses of information

One important application of agency theory is comparing how information is used for managerial incentive purposes relative to how it is used for valuation purposes. Stock price represents the end result of trading decisions investors make based on the information they obtain about the value of the firm. Since stock price directly affects investors' wealth, on the surface it might seem that basing compensation on stock price would be the ideal way to align the interests of managers and shareholders.[46] Indeed, many articles in the business press, consulting newsletters and research journals tout the

---

[45] Obviously, the other division's profits would also be affected by the actions of its own manager. As long as the production function is additively separable, the incentive problems with the two managers can be solved separately, so there is no advantage to explicitly modeling the second manager.

[46] Stock-based compensation means compensation based on the stock price at the end of the period. In a single-period model where the manager has no wealth restrictions, there is no difference between contracts based on end-of-period stock price and contracts based on stock price return. The intercept of the agent's compensation contract is adjusted to offset the differential wealth effects of basing compensation on price versus change in price.

advantages of stock price as a performance measure for exactly this reason. However, agency theory shows that, in general, the way information is aggregated for valuation purposes is not the same way this information would be aggregated for compensation purposes. That is, valuing the firm is not the same as evaluating the contribution of the manager.

The easiest way to see this is to consider the agency models where the outcome x is observable. In this case, the valuation function is trivial: the end-of-period stock price is equal to x (or, more accurately, x minus the agent's compensation). There is no role for other information in the valuation equation. However, from a compensation perspective, there is a role for other information. Specifically, we would like to be able to untangle the effect of the manager's actions from the effect of "other factors" on the outcome x. That is, suppose we write the outcome function as $x = a + \varepsilon$, where a is the agent's action and ε is a random variable that represents other factors that affect the outcome. From a valuation perspective, we care only about the sum of a and ε; however, from a compensation perspective, we care about the individual components. In both the single-action models (see Eq. (8)) and the multi-action models (see Eq. (21)), when x is observable there is still a role for additional performance variables in the contract as long as they are incrementally informative about the agent's actions. These variables need not be incrementally informative about the outcome, just about the actions.

Intuitively, the value of supplementing stock price with other performance measures such as accounting numbers is likely to be higher in single-action models than in multiple action models because in single-action models, basing compensation on accounting numbers has no distortive incentive effects. In a single-action model, both x (stock price) and y_i (accounting earnings) are increasing functions of the agent's action. Therefore, using y_i to increase the agent's action will also increase the outcome x_i. In multiple action settings, it is likely that accounting numbers are insensitive to some actions that would increase the outcome and sensitive to some actions that have no effect on the real outcome. That is, accounting numbers are less congruent than stock price. The principal must trade off the effort distortion that will arise when compensation is based partially on accounting numbers with the risk reduction benefits.

When the outcome x is not observable, the pricing mechanism in the market must aggregate the information signals available to participants. How does the weighting scheme in the valuation function compare to that in an optimal compensation function? If the basic signals are weighted in exactly the same way in the compensation and valuation functions (or are proportional to each other), then the valuation and compensation uses are "identical". Let the basic signals be denoted y_1 through y_n, and suppose the optimal contract is $s = \alpha + \sum_{i=1}^{n} \beta_i y_i$ and the valuation or pricing formula is $p = \gamma_0 + \sum_{i=1}^{n} \gamma_i y_i$. If there is a scale factor k such that $\beta_i = k\gamma_i$ for all i, we can rewrite the optimal

contract as a function of price, p.[47] Specifically, the contract $w = \alpha + k(p - \gamma_0)$ will replicate the original contract based on y_1 through y_n. If this happens, there is no loss to simply basing the agent's compensation on stock price.[48]

To explore this, suppose initially the signals y_i are publicly observed and that a risk neutral valuation formula applies (this is consistent with the principal being risk neutral). With the outcomes and performance measures normally distributed, the pricing function is linear and can be written as[49]

$$p = \gamma_0 + \gamma_1 \tilde{y}_1 + \gamma_2 \tilde{y}_2, \quad \text{where } \gamma_0 = \mathrm{E}(x|\hat{\mathbf{a}}) - \gamma_1 \mathrm{E}(y_1|\hat{\mathbf{a}}) - \gamma_2 \mathrm{E}(y_2|\hat{\mathbf{a}}).$$

This can be rewritten as

$$p = \mathrm{E}(x|\hat{\mathbf{a}}) + \gamma_1 [\tilde{y}_1 - \mathrm{E}(y_1|\hat{\mathbf{a}})] + \gamma_2 [\tilde{y}_2 - \mathrm{E}(y_2|\hat{\mathbf{a}})]. \tag{22}$$

Price depends on both the realization of the performance measures, $\tilde{\mathbf{y}}$, and the conjectured actions, $\hat{\mathbf{a}}$, that market participants expect the agent to take. The realization of the performance measures depends on the agent's actions and on the realizations of the disturbance terms. Note that in pure moral hazard problems, only the disturbance terms are random variables. The conjectured actions depend, of course, on the incentives built into the contract. However, in pure moral hazard agency problems, the market (and the principal) has enough information about the production environment and the incentive contract to be able to correctly conjecture the actions the agent will select. Therefore, the actions the agents will select, $\mathbf{a}$, will exactly equal the actions conjectured by the market, $\hat{\mathbf{a}}$.

As a result, the realizations of the performance measures $\tilde{\mathbf{y}}$ do not cause the market to "learn" anything about the agent's actions. The change in the value of the firm once the market observes the performance measures relates solely to the market updating its assessment of the "disturbance terms" in the signals, $\tilde{\mathbf{y}}$. Therefore, the weighting of the signals in the valuation formula relates solely to the correlation structure between the disturbance terms in $\tilde{\mathbf{y}}$ and the random component of the outcome, x. In fact, the weights in the valuation formula are identical to the slope coefficients from a multiple regression of x on y_1 and y_2. The slope coefficients depend only on the covariance matrix of the disturbance terms $(\tilde{e}_x, \tilde{e}_{y_1}, \tilde{e}_{y_2})$. For example, if $\tilde{e}_{y_i}$ and $\tilde{e}_{y_2}$ are uncorrelated, the slope coefficients in the valuation function are simply $\gamma_i = \mathrm{cov}(\tilde{e}_x, \tilde{e}_{y_i})/\mathrm{var}(\tilde{e}_{y_i})$. At the extreme, if the disturbance term in a signal is uncorrelated with the disturbance term in the outcome, the signal will receive *zero* weight for valuation purposes. Note that for valuation purposes, the valuation weights do not depend on the

---

[47] Note that for this to happen, the same basic information signals used to form price must also be available for contracting purposes. Moreover, there cannot be other sources of "noise" in price.

[48] See Feltham and Xie (1994) for some examples.

[49] The stock price should also reflect the compensation paid to the agent, so that it is more accurate to write $p(\tilde{\mathbf{y}}) = \mathrm{E}(x|\tilde{\mathbf{y}}, \hat{\mathbf{a}}) - s(\tilde{\mathbf{y}})$. This makes the math a little more complicated to do, but does not change any of the insights.

sensitivities of the performance measures to the agent's actions. All this is anticipated by the market and reflected in the intercept of the valuation formula.

In contrast, the sensitivities of the signals to the actions play a critical role in the weights in the compensation function. For example, in a single action, we saw (in Eq. (11)) that the relative weights are a function of the performance measures' sensitivity to the action, their variances, and their covariance with the other. Note that the correlation between a performance measure and the outcome is not relevant to how the performance measure is used in the contract. The contrast can also be clearly seen in the situation where there are multi-actions and the agent is risk neutral. In this case, Eq. (18) showed that the signals are weighted to maximize the alignment between the vector of sensitivities of the performance measures with the vector of sensitivities of the outcome to the agent's actions. The correlation structure of the disturbance terms is irrelevant. That is, the important role of the performance measures in the compensation contract is *congruity*, which relates to the sensitivities of the measures. In contrast, the role of the performance measures in the valuation formula is *correlation*, which relates to the disturbance terms.

When the agent is risk averse, the variances of the disturbance terms are also important.[50] However, here variation that is "real" from the perspective of valuing the firm can be noise from the perspective of evaluating the agent's actions. For example, suppose the outcome of the firm can be separated into two components: $x = x_1 + x_2$, where $x_i = b_i a_i + \theta_i$. Note that the disturbance term θ_i is a real cash flow, just not one related to the agent's action. Therefore, the θ_i term is relevant from a valuation perspective, but is noise from a compensation perspective. The observed variables measure the underlying cash flow with noise: $y_i = x_i + \varepsilon_i = b_i a_i + (\theta_i + \varepsilon_i)$. For convenience assume all the θ_i and ε_i are independent. Note that ε_i is noise from both a valuation and a compensation perspective, whereas θ_i is noise only from a compensation perspective.

If the contract can be based on the variables y_1 and y_2, the contract will make the weight on each variable a decreasing function of its variance. That is, the weight on y_i is a decreasing function of both the variance of θ_i and the variance of ε_i. However, in the valuation formula in Eq. (22) above, the weight on signal y_i is

$$\gamma_i = \frac{\text{cov}(x, y_i)}{\text{var}(y_i)} = \frac{\text{cov}(\theta_1 + \theta_2, \theta_i + \varepsilon_i)}{\text{var}(\theta_i + \varepsilon_i)} = \frac{\text{var}(\theta_i)}{\text{var}(\theta_i) + \text{var}(\varepsilon_i)}$$
$$= \frac{1}{1 + \text{var}(\varepsilon_i)/\text{var}(\theta_i)}. \qquad (23)$$

---

[50] See Paul (1992) for additional discussion.

This is the classic "errors in variables" formula in econometrics. Note that the slope coefficient in the valuation formula is decreasing in $var(\varepsilon_i)$, just as it is in the compensation formula.

However, in contrast to the weights in the compensation function, the slope coefficient in the valuation formula is *increasing* in $var(\theta_i)$. That is, the more "real" variation there is in the performance measure, the more sensitive price is to the realization of the measure. Unfortunately, none of this "real" variation is related to the agent's action, so the pricing formula "over-weights" those performance measures with high levels of real variation compared to "measurement error". This distortion in the weights has real incentive effects on the agent's action choices. In particular, if the agent is compensated solely on the basis of price (so the weights in the valuation function become the incentive weights for the agent), the agent's effort gets skewed toward those actions that have higher "real" variability than would be the case if the compensation function was based on the underlying variables y_1 and y_2.

As discussed above, one important reason for the stark contrast between information aggregation for incentive and valuation purposes is that for incentive purposes, the sensitivities are important, whereas for valuation purposes they are not. This result occurs because the agent's action is selected based on the same information that investors have. Therefore, investors are able to perfectly conjecture what actions will be selected and incorporate this into the stock price "intercept". The signal realization does not help the market update its assessment of the "action" component of the outcome; it only helps the market predict what the "disturbance" term of the outcome will be.

However, in more general models the market will be uncertain about the agent's actions (perhaps because the agent has different information about the profitability of different actions) or the agent's talents. In this case, the market will use the realizations of the performance measures to update their assessments of *both* components of the outcome. Intuitively, this may bring the weights placed on the signal for valuation purposes closer to the weights placed for compensation (or salary adjustment) purposes. To date, little work has been done on these kinds of private information models (or more generally, stochastic actions or talent) to analyze these issues.[51] In Section 4, I discuss the literature on private information models.

3.3.6. Other issues in the use of information for valuation versus stewardship purposes

One important role that stock price plays is that of aggregating diverse pieces of information relevant for estimating firm value. In many instances, the

---

[51] See Baker (1992) and Bushman et al. (2000) for additional analysis.

individual information variables collected by investors are not publicly observable or are not in a form that makes them easy to include in a compensation contract. Since stock price is a contractible variable, it can serve as a valuable proxy for the underlying information signals. A number of papers have developed and analyzed noisy rational expectations models in which investors observe private signals and then trade to an equilibrium.[52] In addition to the features discussed above, in a noisy rational expectations model, the equilibrium stock price also contains "noise" (by definition). As a result there is a role for other information to play in both valuation (investors condition their demand based on price and their own private information) and compensation (if other performance variables can be found that remove some of the noise in price).

It is also possible that the contract can be based on variables that are unobserved by the market. That is, the board of directors may have access to inside information that is too costly to (or impossible to) credibly disclose to the market. In such a setting, there are obviously gains to using this additional information as an input into the compensation equation.

Another important dimension to understanding the costs and benefits of using stock-based compensation relates to how forward-looking stock prices are relative to other performance variables. Stock price is commonly thought to be extremely forward looking; at the extreme, it is thought to represent the present value of all future cash flows based on all information available to investors at the time. This forward-looking property of stock prices is generally viewed to be a beneficial feature of using stock price as an incentive measure, particularly when there are concerns about the decision-making horizon of managers. However, Barclay et al. (2000) argue that this forward-looking property of stock prices can also be a negative feature for incentive purposes. In particular, stock prices will reflect the market's expectations about not only the future period implications of managerial actions already taken, but also the expected impact of *future* period actions. Therefore, stock-based compensation rewards managers in advance for profitable actions that are expected but have not yet been delivered. If managers can leave the firm and contracting frictions prevent shareholders from recouping compensation that has already been paid to the manager, then compensating managers in advance (using stock price) can make shareholders worse off than delaying compensation until the performance has been delivered (using, say, earnings). Multi-period aspects of compensation and performance measures are discussed in more detail in Section 6.

Finally, research into understanding the costs and benefits of using stock-based compensation should incorporate the possibility that the manager can trade in the stock market on his own behalf. Depending on the nature of these

---

[52] See Bushman and Indjejikian (1993a, b), Kim and Suh (1993), and Feltham and Wu (2000).

trading opportunities, the observability of the manager's trading decisions, and the size of his outside wealth, the ability of the manager to trade can have a significant impact on the desirability of using stock-based compensation. At the one extreme, these trading opportunities could allow the manager to undo any incentives the principal attempts to impose through the compensation contract. The manager's incentives to collect private information and to publicly disclose information to the market can also be greatly affected by his trading opportunities (i.e., insider trading). See Baiman and Verrecchia (1995) and Stocken and Verrecchia (1999) for analyses of these issues. I discuss private information and communication in agency models in more depth in the next section.

4. Private information and communication

In this section, I discuss research that extends the basic agency model by allowing one party to have private information. For the most part, researchers have assumed that the agent is the party who obtains private information.[53] This could be information he receives prior to entering the agency relationship (information about his "type", e.g., his skill, expertise, degree of risk aversion, or minimal acceptable level of utility). In other instances, this private information is acquired once he is on the job. That is, by virtue of being closer to the production process, the customers, etc., it is natural to think that the agent would become better informed about the "operational" aspects of the firm. Similarly, if the agent proposes new projects or investments, it will generally be the case that he has superior information about the expected profitability or the timing of the payoffs than does the principal.

For concreteness, let m be the information signal observed by the agent. As before, let x and $\mathbf{y}$ denote the outcome and any additional signals that are publicly observed at the end of the period. Let the a priori probability density function of the signals be $g(m)$. Once a signal is received, the density function of the outcome and other performance measures is updated to be $h(x, y | a, m)$.

There are generally three dimensions that researchers must consider in formulating a private information model. First, when does the agent receive the information—before he signs the contract, after he signs but before he selects his actions, or after he selects his actions? The second dimension is whether the agent can leave the firm after observing the information signal. The third is

---

[53] There are considerably fewer papers which model the principal as having superior information. It would be plausible to expect the principal to have superior information about "strategic" variables than would agents below the top management level. The principal would likely also have superior information about the activities of other subunits of the firm, and how the activities across units need to be coordinated.

whether the agent is allowed to communicate this signal (perhaps untruthfully) to the principal.

We begin with the case where the agent receives private information after he signs the contract but before he selects his action, he cannot leave after observing the signal, and he cannot communicate the signal to the principal. When the agent receives private information, he can tailor his actions to the specific information he has received. Let $a(m)$ denote the agent's action strategy as a function of the information signal observed. Naturally, this strategy will be influenced by the incentives set up by the principal. The agent's compensation, $s(x, \mathbf{y})$ cannot depend directly on m because m is unobservable. Nevertheless, if m affects the distribution function of x and y, the agent's optimal action is likely to change as a function of the signal observed. The principal's problem is now

$$\max_{s(x,y),a(m)} \mathrm{E}_{x,y,m}[G[x - s(x, \mathbf{y})]|a(m)]]$$

$$\text{subject to } \mathrm{E}_{x,y,m}[U[s(x, \mathbf{y})|a(m)] - V[a(m)]] \geqslant \underline{H},$$

$a(m)$ is the a that maximizes $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y})|a] - V(a)$

for each signal m.

The major difference between this and the symmetric information formulation is that there is a *set* of incentive compatibility constraints— one for each possible signal. The calculation of the principal's and agent's expected utilities is also more complicated since the agent's action will generally depend on the signal received. For example, the ex ante expected outcome is derived by considering the probability density function of the signals, $g(m)$, the agent's optimal action if he observes signal m, $a(m)$, and the expected outcome given the signal and the optimal action, $\mathrm{E}[x|a(m), m]$.

When the agent cannot leave the firm, the principal's maximization problem only requires the principal to meet the minimal acceptable utility constraint in expectation. Even if the principal is able to design the contract to make this constraint binding in expectation, it is unlikely it will be binding for each realization of the signal.[54] Therefore, there are likely to be some realizations for m such that the agent no longer believes he will receive at least as much expected utility from continuing to work for the principal under the existing contract as his other employment opportunities provide. If the agent has the

---

[54] As in the symmetric information case, the principal can hold the agent to the minimal acceptable utility level in expectation if there are no constraints on the minimal payment that can be made to the agent or if the agent's utility function is unbounded below as his payment approaches any minimal allowable payment.

ability to leave the firm, it would be in his best interests to do so after observing "unfavorable" realizations of m. If the principal wants to prevent the agent from leaving under any circumstances, he must offer a contract that is attractive for all realizations, not just in expectation.[55] In such a setting the minimal acceptable utility constraint in the contract is replaced by a set of constraints of the form

$$E_{x,y|m}[U[s(x,\mathbf{y})|a(m)] - V[a(m)]] \geqslant \underline{H} \quad \text{for all } m.$$

In general, the principal will be worse off when the agent has the opportunity to leave. In particular, if the agent earns the minimal acceptable level of utility for the "worst" realization of m, then he earns in excess of the minimal acceptable level of utility for more "favorable" realizations.[56] The agent's ex ante expected utility therefore exceeds the minimal acceptable level. This excess level of utility is termed an "information rent"; the agent's power in the relationship is increased by his superior information about m.

The principal can potentially reduce the information rent by having the agent communicate the information signal to the principal. If the reporting is exogenously constrained to be truthful, the principal can include the report in the contract as a publicly observed signal. In particular, the principal can tailor the compensation contract to the specific information signal received. For example, if the signal revealed that the marginal productivity of effort was high (low) the principal might increase (decrease) the sensitivity of the agent's compensation to the outcome in order to motivate him to provide more effort. Similarly, if the signal was informative about the expected value of output, the principal might "subtract out" the expected output that was due to the signal m in order to better focus on the part of output that is due to the agent's effort.

Of course, it is unlikely to be the case that there is an exogenous guarantee that the agent's reports will be truthful. The principal must anticipate that the agent has an incentive to misreport the signal he saw in order to receive a more "favorable" compensation contract. We model this by denoting the agent's report as communicating a message (or report), $\hat{m}(m)$, where $\hat{m}$ is the message the agent sends after seeing the information signal m. We then augment the principal's problem with a group of self-selection constraints for the agent's

---

[55] Implicitly, these models assume that if the first agent leaves, an "otherwise identical" agent must be hired and that this second agent would observe the same information signal the first agent did. Since there is no advantage to hiring the second agent relative to retaining the first, the models assume the first agent is retained. It is easy to modify this to include situations where the principal chooses to "abandon" the firm or declare bankruptcy for some realizations of m.

[56] It is also possible to model the situation where the utility level the agent can exceed outside the firm is lower than $\underline{H}$ if he leaves in the "middle" of the period.

message strategy. The principal's problem becomes

$$\max_{s(x,\mathbf{y},\hat{m}),a(m),\hat{m}(m)} \mathrm{E}_{x,y,m}[G[x - s(x, \mathbf{y}, \hat{m})]|a(m)]]$$

subject to $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, \hat{m})|a(m)] - V[a(m)]] \geqslant \underline{H}$ for all m,

$a(m)$ is the a that maximizes $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, \hat{m})|a] - V(a)$
for each signal m,

$\hat{m}(m)$ is the $\hat{m}$ that maximizes $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, \hat{m})|a] - V(a)$ $\forall$ signals m.

Note that contract depends on the observable variables x and y as well as the message issued by the agent. One interpretation is that the principal pre-commits to this compensation function and then the agent issues his message. An equivalent interpretation is that the principal offers a menu of contracts $\{s(x, \mathbf{y})\}$ that the agent can select from. The agent then selects the contract from that menu that maximizes his expected utility given the information signal he observed. From a modeling perspective, selecting a contract from this menu based on the information signal observed is equivalent to communicating this information.

The first set of constraints are the minimal acceptable utility constraints, the second set of constraints are the incentive compatibility constraints on the agent's effort, and the third set of constraints are the incentive compatibility constraints on the agent's reporting strategy.[57] Of course, the real information signal m is what the agent uses to update his expected utility conditional upon its receipt. The message only affects the contract the agent receives.

When the principal must decide what message strategy to motivate, the principal's problem becomes very difficult to solve because the class of message strategies to search over is large and "messy". Fortunately (and also unfortunately, as I discuss in a later section), researchers have found that the revelation principle allows them to greatly simply their problem formulation and solution. The revelation principle was developed in the mechanism design literature (see Myerson, 1979), and it applies when agents receive private information, agents have the ability to transmit their information in its full dimensionality (i.e., if they observe two signals they are permitted to transmit a two-dimensional message and are not forced to aggregate the signals into a one-dimensional message), and the principal is able to credibly commit to how the information will be used.

The revelation principle states that any proposed mechanism that involves nontruthful reporting by the agent can be duplicated or beaten *in terms of*

---

[57] A better way to express the incentive compatibility constraints is that the joint strategy (action strategy, reporting strategy) is in the agent's best interests. Modeling them separately as we have done in the text leaves open the possibility that each strategy is optimal given the other, but the strategy pair is merely a local maxima for the agent, not a global one given the contract by the principal.

expected utilities by an equilibrium mechanism in which truthful reporting is induced. Similarly, any multi-stage process (the agent submits a tentative message, the principal makes a counter-offer, the agent submits a revised message, etc.) can be duplicated or beaten by a single-stage process in which the agent submits the truth. It is important to recognize that the revelation principle *does not* say that truth telling comes at zero cost. On the contrary, the principal must design the contract to induce the agent to tell the truth. In general, this will force the principal to pre-commit to "under-utilizing" the information. That is, the cost of inducing the agent to tell the truth is that the principal cannot use the information as fully as he would if the truthful message did not have to be motivated. In fact, in some extreme cases the principal must promise to not use the information at all in order to induce the agent to report honestly. The revelation principle merely states that the cost (broadly defined) of motivating the truth is no greater than the cost of motivating a nontruthful reporting strategy.

Essentially, the revelation principle works by collapsing any proposed misreporting into the actual contract. That is, suppose the principal is considering offering the contract $s^1(x, y, \hat{m})$, and that this contract will motivate the agent to select an action strategy $\mathbf{a}(m)$ and a reporting strategy $m^1(m)$. Now consider a different contract in which the last argument in the contract is replaced with the message strategy that was optimal for the agent under the first contract. That is, let $s^2(x, y, \hat{m}) = s^1(x, y, \hat{m}) \circ m^1(m)$. Then the agent's optimal action strategy is the same as before, and if $m^1(m)$ was his best report strategy under contract 1, then m is his best action strategy under contract 2. The expected utilities are therefore equivalent under contract 1 (which potentially involves misreporting) and contract 2 (which involves truth telling).

The revelation principle has been an extremely valuable methodological tool because it greatly reduces the number of alternative reporting strategies researchers must consider as possible equilibria in their models. It implies researchers can safely confine their attention to equilibria which motivate truthful reporting. Applying the revelation principle to the principal's problem above allows us to write it as

$$\max_{s(x,\mathbf{y},m),a(m),m(m)} \mathrm{E}_{x,y,m}[G[x - s(x, \mathbf{y}, m)]|a(m)]]$$

subject to $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, m)|a(m)] - V[a(m)]] \geqslant \underline{H}$ for all m,

$\quad a(m)$ is the a that maximizes $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, m)|a] - V(a)$
$\quad\quad$ for each signal m,

$\quad m(m)$ is the $\hat{m}(m)$ that maximizes $\mathrm{E}_{x,y|m}[U[s(x, \mathbf{y}, \hat{m})|a] - V(a) \ \forall$ signal m.

Note that the agent's report has been replaced by the true signal in the contract. However, the contract must ensure that truth telling is the best reporting strategy for the agent.

4.1. Application to capital budgeting

One important application of agency theory models with private information relates to budgeting procedures in general and capital budgeting in particular[58] (see Antle and Fellingham (1997) for recent review.) Antle and Fellingham (1997) analyze a capital budgeting model (similar to the one analyzed by Antle and Eppen, 1985) in which the agent receives private information about the profitability of investment. Equivalently, and we will use this interpretation, we can think of the agent as learning the minimal investment amount that will yield a given end-of-period cash flow x. In particular, to produce an end-of-period cash flow of x, the minimal investment is mx. The agent does not possess the capital to fund the investment on his own. Instead, the capital must be supplied by the principal. Let z be the amount of resources transferred from the principal to the agent. The amount z is to cover the investment as well as the agent's compensation. The agent can choose to divert a portion of z for his own purposes, which we interpret as the agent consuming "slack".[59]

It is common in the capital budgeting area to work with a discrete number of information signals. Accordingly, we order the signals from low cost (most favorable) to high cost (least favorable): $m_1 < m_2 < \cdots < m_n$. The ex ante probability that the cost signal is m_i is denoted g_i.

The principal is assumed to be risk neutral, and his utility is defined over the end-of-period cash flow minus the resources transferred to the agent, $x - z$. The agent's utility is defined over the amount of resources he receives at the beginning of the period minus the amount he invests in the production process, $z - m_i x$. This amount consists of the agent's compensation plus the amount of slack he consumes. The agent's reservation level of utility is $\underline{H}$. Finally, the agent can leave after observing the cost realization. Therefore, the contract must be attractive enough to keep the agent even if the cost realization is "really bad".

There are several features of these capital budgeting models that are different from the moral hazard models. First, note that both the principal and the agent are risk neutral. In private information models, a primary source of the principal's welfare loss is due to the information rent possessed by the agent. In general, this information rent exists even if the agent is risk neutral. Unlike moral hazard problems, the principal cannot simply eliminate the information

---

[58] See Kirby et al. (1991) for an application to participative budgeting, Reichelstein (1993) for an application to forecasting costs for government contracts, and Antle and Fellingham (1997) for a review of capital budgeting papers.

[59] To increase the comparability with the moral hazard models discussed earlier, we can interpret the model as the agent being able to substitute capital for labor. If the agent is able to convince the principal to provide more capital, he does not have to work as hard to achieve a given level of output.

rent by "selling the firm" to the agent because the agent has superior information about what the firm is "worth". For this reason, many research papers examine private information issues using risk neutral agents, so they can avoid the complexities that result when risk aversion and risk sharing issues are also present.

Second, there is a communication feature of the model. The agent issues a report of the cost, $\hat{m}$, and the principal will make the inputs provided, z, and the output required, x, dependent on the report. The principal pre-commits to how the output and resources provided will depend on the report.

Finally, although this is not a general feature of private information models, once the agent observes the information signal, he has perfect information about the cost; he faces no residual uncertainty. On the one hand, this makes the risk neutrality of the agent of less concern. In general, the lack of residual uncertainty allows the principal to infer more from the outcome than he otherwise might be able to do. In particular, it will prevent the agent from overstating the amount of output he can produce from a given level of resources. However, other features of the model (especially the linear production technology combined with a bounded level of output) work the other way. That is, the exact same level of output will be produced for many different information signals; this reduces the ability of the principal to use the realized output level to infer what the agent saw.

4.1.1. First-best solution

Suppose the cost signal was publicly observed (or was automatically communicated truthfully to the principal). The principal's problem is then

$$\underset{x_i, z_i}{\text{maximize}} \sum_{i=1}^{N} (x_i - z_i) g_i$$

$$\text{subject to} \quad z_i - m_i x_i \geqslant \underline{H} \quad \text{for all } i = 1, \ldots, N,$$

$$0 \leqslant x_i \leqslant x_{\max},$$

$$0 \leqslant z_i.$$

The principal maximizes the expected value of net profits. Note that we have normalized the per unit value of output to be 1.0. The first constraint is that the agent receives an acceptable level of utility regardless of the signal observed. The second constraint places bounds on the allowable output. The linearity of the production technology requires bounds in order for a solution to exist. Clearly, we could imagine modifying the model to a concave production technology with no exogenous bounds on output. The last constraint requires the amount of resources transferred to the agent to be nonnegative, i.e., the agent does not have sufficient wealth to provide any of the funding.

Since the principal's problem has been structured to be a linear programming problem, the solution is straightforward. For each cost realization, the agent receives just enough resources to produce the desired output and to cover his reservation utility. That is, $z_i = m_i x_i + \underline{H}$. Substituting gives

$$\underset{x_i}{\text{maximize}} \sum_{i=1}^{N} (x_i - m_i x_i - \underline{H}) g_i = \underset{x_i}{\text{maximize}} \sum_{i=1}^{N} (1 - m_i) x_i g_i - \underline{H}.$$

Note that the objective function is a linear function of x_i, so the solution is to choose

$$x_i = \begin{cases} x_{\max} & \text{if } m_i \leqslant 1, \\ 0 & \text{if } m_i > 1. \end{cases}$$

That is, we produce the maximal possible output ($x_{\max}$) if the per unit cost of output is less than the per unit value of output (i.e., 1.0), and the minimal possible output (zero) is produced if the per unit level of output is greater than its per unit value. Finally, the resources allocated are

$$z_i = \begin{cases} m_i x_{\max} + \underline{H} & \text{for } m_i \leqslant 1, \\ \underline{H} & \text{for } m_i > 1. \end{cases}$$

For each signal, the agent receives enough compensation to cover the desired level of investment plus his reservation level of utility. The agent receives zero slack.

4.1.2. Reporting incentives

Suppose the agent is offered this first-best contract. If he privately observes the cost realization, is it in his best interests to report it truthfully? If the agent observes signal m_i, but reports a cost m_j, he will receive resources and an output corresponding to signal j: $z_j = m_j x_j + \underline{H}$. His problem is to

$$\underset{m_j}{\text{maximize}} \, (m_j x_j + \underline{H}) - m_i x_j = \underset{m_j}{\text{maximize}} \, (m_j - m_i) x_j + \underline{H}.$$

The agent will therefore choose m_j to maximize $m_j x_j$. That is, he will overstate the cost to get more resources than he really needs. However, there is a limit to how much he will overstate. If he says the cost is too high (over 1) the principal will not give the agent any resources. Therefore, the agent will report a cost of slightly under 1 for all $m_i < 1$, and if he observes a cost realization above 1, he has no incentive to lie, because the project will not be funded anyhow. As discussed above, the principal does not have to worry about the agent understating his costs because if he does so, he will not receive enough resources to produce the required output, so a lie in this direction will always be discovered. Note that the principal's ability to forego the project after

observing the agent's report provides some discipline on the agent's incentive to over-report the cost.

4.1.3. Principal's problem when the agent reports strategically

When the principal recognizes the agent's ability to manipulate the report, we model the principal as choosing a menu of (x_i, z_i) combinations that the agent can choose from conditional upon observing a cost signal m_i. The principal's problem is

$$\underset{x_i, z_i}{\text{maximize}} \sum_{i=1}^{N} (x_i - z_i) g_i \tag{24}$$

subject to $z_i - m_i x_i \geqslant \underline{H}$ for all $i = 1, \ldots, N$, $\qquad$ (24a)

$\qquad z_i - m_i x_i \geqslant z_j - m_i x_j$ for all i, j, $\qquad$ (24b)

$\qquad 0 \leqslant x_i \leqslant x_{\max}$, $\qquad$ (24c)

$\qquad 0 \leqslant z_i$. $\qquad$ (24d)

This problem differs from the first best by the addition of the second set of constraints in Eq. (24b). These "truth-telling" constraints require that when the agent observes cost signal m_i, he is better off choosing the (output, resource) combination (x_i, z_i) than any other combination (x_j, z_j). Note that the agent's report does not affect the underlying cost. That is, after observing signal m_i, this is the real cost of whatever output level the agent claims to be able to achieve. That is, the true m_i is what multiples the value of x on each side of the constraints in Eq. (24b).

We can use the truth-telling constraints alone to derive a lot about the solution to the principal's problem. In particular, these constraints imply (see Antle and Fellingham (1997) for a proof):

(i) The output, x_i, is weakly decreasing in i.
(ii) The resources provided, z_i, are weakly decreasing in i.
(iii) The agent's slack, $z_i - m_i x_i$, is weakly decreasing in i, and is strictly decreasing for ranges of j where $x_j > 0$.

The constraints in the principal's problem can also be simplifying by noting that

(iv) If the acceptable utility constraint in (24b) holds for the worst cost realization (m_N), it will hold for all others. This also implies $z_N - m_N x_N = \underline{H}$.
(v) If the incentive compatibility constraint in (24c) for m_i is met for the next higher cost realization (m_{i+1}), it will be met for all others.

These are common properties of the solution to the communication in budgeting models. They greatly reduce the number of constraints in the model, thereby making it much easier to solve.

The solution to the principal's problem has a simple structure. First, there is a threshold level of cost that determines whether any output is funded. That is, there is a threshold level of j such that

$$x_j = \begin{cases} x_{\max} & \text{for } m_j \leqslant \text{threshold,} \\ 0 & \text{for } m_j > \text{threshold.} \end{cases}$$

The resources provided are

$$z_j = \begin{cases} m_{\text{threshold}} x_{\max} & \text{for } m_j \leqslant \text{threshold,} \\ 0 & \text{for } m_j > \text{threshold.} \end{cases}$$

Note that the agent receives excess resources for all cost realizations below the threshold. Therefore, "slack" arises endogenously as part of the optimal solution. For these signals, the principal is precommitting to allow the agent to keep this slack, and not renegotiate the contract after the agent makes this report. This pre-commitment to underutilize the information is what gives the agent incentive to reveal it truthfully.

The only remaining choice is the threshold level of cost for determining whether to fund the project or not. Interestingly, the optimal cutoff point is below 1.0. That is, the optimal solution involves the principal turning down some projects that are profitable. One interpretation of this is capital rationing. In contrast to other explanations of capital rationing that rely on exogenous limits on the amount of capital that is available for funding projects, here capital rationing arises as part of the optimal equilibrium. The principal is willing to do this in order to reduce the slack the agent is able to create. When the principal changes the threshold level of m, he affects the amount of slack the agent is able to create for all lower levels of m. Therefore, by reducing the threshold below 1.0, the principal is foregoing production in some states in order to reduce the slack the agent consumes in the other states.

4.2. Role of additional (ex post) performance measures in private information models

In contrast to the agency models that study pure moral hazard problems, little has been done to analyze the role of (or weights assigned to) additional performance measures observed at the end of the period in private information models. Work to date (as well as speculation on my part) suggests there are several important differences between the factors that drive the optimal weighting of performance measures in the two settings.

First, it is unlikely that the variance of a signal can be viewed as a measure of the noisiness of the signal. When the agent's actions depend on the realization of the private information signal, then the outcome (x) as well as ex post performance measures (y) will have a stochastic component that is "real" (i.e., due to the stochastic nature of the agent's actions) and a stochastic component that is "noise" (is unrelated to the agent's actions or the private information observed ex ante). In this setting, performance variables that do not exhibit much variation are unlikely to be valuable because they do not reflect the real underlying variation in the actions. At the other extreme, performance measures that exhibit too much variation are unlikely to be valuable because they are too affected by noise. The challenge to the principal is to determine the performance measure's conditional variance, that is the variance conditional upon the agent's action and private information signal. Estimating this variance is also a challenge for researchers attempting to test theories regarding how the "noisiness" of a signal affects the contract parameters.[60]

Second, the notion of congruity must be adapted to reflect the possibility that the sensitivity to the outcome or a performance measure may be stochastic. That is, a performance measure must not only be congruent across actions but also across states (or signals).[61] To see this, consider a single-action model in which the agent observes a signal prior to selecting his actions. Suppose the outcome is not observable, and the principal must choose between basing compensation on either performance measure y_1 or performance measure y_2. On average, both y_1 and y_2 are equally sensitive to the agent's action, and this average sensitivity is the same as the sensitivity of the outcome to the agent's action. In this setting, if the agent did not receive private information before taking his action the principal would be indifferent regarding which performance measure to use, ceteris paribus. This is not true when the agent receives private information prior to his action choice. For example, suppose that conditional upon receiving a private information signal, the sensitivity of signal 1 is high (low) when the sensitivity of the real outcome is high (low), while the sensitivity of signal 2 is unrelated to the sensitivity of the real outcome. Clearly the principal would prefer to contract using performance measure 1 than 2. Therefore the correlation between the sensitivity of the performance measure and the sensitivity of the outcome is likely to be important.

---

[60] Studies which have used the relative performance evaluation hypothesis to extract the market component of or industry component of returns can be viewed as attempting to estimate the conditional variance (i.e., the variance of the firm-specific component of returns); see Antle and Smith (1986) and Janakiraman et al. (1992) for examples. Sloan (1993) estimates the conditional variance of accounting performance measures by extracting the part of the signal that is related to the firm-specific component of stock price.

[61] See Baker (1992) and Bushman et al. (2000) for additional analysis.

Related to this is the notion that the performance measure should be informative about the agent's action strategy. Simply knowing that the agent's effort level was "low" is not enough; it is important for the principal to be able to tell whether this was because the agent goofed off or because the agent received information that indicated that effort was not going to be very productive. Ex post performance measures can therefore be valuable even if they do not tell you anything directly about the agent's action, but instead tell you about the private signal. Moreover, ex post performance measures can be used to discipline the agent's report. That is, they can be used to reduce the agent's information rent. Therefore, in these models, accounting numbers can have a value even if the only role they play is to confirm prior disclosures. At the extreme, the equilibrium could be such that the accounting numbers released at the end of the period have no "information content", i.e., market participants do not react to the release of the accounting numbers and the accounting number does not appear to affect the agent's compensation. Nevertheless it is the threat that the accounting number will differ from the agent's earlier disclosure that makes the agent's earlier disclosure credible in the first place. While I have mainly concentrated on the agent disclosing information received prior to selecting his action, this clearly applies also to private information received after the action is taken, but before the outcome and/or performance signals are generated.

Finally, competition among agents can be valuable in private information settings. In particular, competition among informed agents can be used to reduce the information rent the principal must give up. In contrast, in pure moral hazard models, competition has no direct benefit; instead the benefits from making an agent's compensation depend on the performance of others arise because this allows the principal to filter out "systematic risk" (i.e., the effect of shock terms that are common across agents) from the agent's compensation.[62]

4.3. Incentive problems and the charge for capital

In this section, I discuss the application of private information models to issues of whether and how much managers should be charged for the capital they use. While this is obviously important in its own right, it is also a fundamental part of the construction of financial performance measures like residual income and economic value added (EVA), in cost allocation, and in transfer pricing. In particular, EVA theorists and most accounting textbooks

---

[62] In a more general equilibrium model, competition among agents is likely to reduce the agent's bargaining power with both this principal and other potential employers, which could affect the level of the agent's reservation utility. However, holding the reservation level of utility fixed, competition has no direct benefit in pure moral hazard problems.

generally claim that the cost of capital charged to managers should be identical to the firm's cost of capital. Recent theoretical research by Rogerson (1997), Reichelstein (1997, 2000), and Dutta and Reichelstein (1999) also supports the idea that the charge for capital should be the owner's cost of capital. However, other agency models imply that the charge for capital should not be the same as the owner's cost of capital. In particular, these papers suggest that the optimal charge for capital is often higher than the owner's cost of capital. In this section, I discuss some of these findings.

For example, we can recast the Antle and Fellingham (1997) paper as a cost of capital paper. Specifically, if $m_i x$ is the amount of investment needed to yield an end-of-period cash flow of x, the rate of return on investment, r_i, satisfies $m_i(1 + r_i) = 1$, or $r_i = 1/m_i - 1 = (1 - m_i)/m_i$. Their result that the optimal solution is to establish a threshold level of cost, m^*, and invest only if $m_i < m^*$ is equivalent to establishing a hurdle rate of return r^* and investing only if the rate of return, r_i, exceeds this threshold rate of return, i.e., invest if and only if $r_i > r^* = 1/m* - 1$. Moreover, for those states where the agent's investment project is accepted, he is charged r^* per dollar of capital he receives from the principal. Since the threshold level of cost is generally less than 1.0, the cost of capital charged for the agent is positive. This occurs even though the cost of capital for the principal has been normalized to zero. Therefore, the cost of capital charged to the agent is higher than the external cost of capital to the firm (principal).

As discussed in Antle and Fellingham (1995), when additional information is available at the end of the period, this can potentially reduce the information rents obtained by the agent and result in more efficient production. They also show that it is not always the case that this implies a lower cost of capital charged to the agent. It is not clear what features of their model lead to this result, or what conditions would ensure that better information would lower the cost of capital (or whether this is even the right question). I suspect that in models with continuous investment amounts, concave production functions, and "smoother" information systems, it will be easier to establish the nature of the link between quality of information and cost of capital.

Next, I consider a model that is closer in spirit to those analyzed in Rogerson (1997) and Reichelstein (1997, 2000). Here, the issue is not one of reducing the agent's information rents per se, but instead is that of taking into consideration the interplay between investment incentives and other incentive problems. Specifically, I derive conditions where the investment incentives can be "uncoupled" from the other incentives so that the correct charge for capital is the principal's cost of capital. In this case, EVA or residual income is the "right" performance measure. However, I also show conditions where this result does not hold, so that the capital charge for investment must take into consideration the nature and severity of the other incentive problems in the model. In this situation, EVA does not give the correct charge for capital.

Assume the agent is responsible for "productive effort" (*a*) and a capital investment (*I*). The productive effort is similar to the effort described in previous sections; it will generate a cash flow, $\tilde{x}_a$, at the end of the period, which is normally distributed with an expected value of ba, and a variance of $\text{var}(x_a)$. That is, $\tilde{x}_a = ba + e_a$, where e_a is normally distributed. The agent bears a personal cost, $C(a) = 0.5a^2$ for the first type of effort.

If the agent invests *I* in the capital investment at the beginning of the period, it will yield a normally distributed cash flow of $\tilde{x}_I$, whose expected value increases (at a decreasing rate) with *I*, and whose variance is independent of the magnitude of *I*.[63] Specifically, let $\tilde{x}_I = \text{E}(x_I|I, m) + e_I$ where *m* is a (possibly private) information signal about the productivity of investment. For convenience, $\tilde{x}_a$ and $\tilde{x}_I$ are independently distributed. In contrast to the models in Rogerson (1997) and Reichelstein (1997, 2000), the agent has a nonpecuniary return $V(I)$ associated with the magnitude of the investment. For now, we will not specify the form of $V(I)$, but it could be increasing (he gets more prestige from bigger investments, or he can consume "slack") or $V(I)$ could be decreasing (bigger projects require more work to implement).

The principal discounts money using an interest rate of *r* per period. For purposes of calculating present values, assume the cash flows at the beginning and end of the period are one period apart. Let $\tilde{X} = \tilde{x}_a + \tilde{x}_I$ denote the total cash flow (gross of the agent's compensation) at the end of the period. If the agent is paid an amount *s* at the end of the first period, the risk neutral principal's utility (as assessed at the beginning of period 1) is $\text{E}((\tilde{X} - s)/(1 + r) - I)$. The first term is the present value of the net cash flow (return from effort and short-run investment minus compensation to the agent) that occurs at the end of the first period. The second term is the investment made at the beginning of the first period.

The agent is weakly risk averse, with a negative exponential utility function. If he is paid *s* at the end of the first period, the certainty equivalent of his

---

[63] When the variance of the outcome depends on the agent's investment decision, the principal and agent may disagree about how to trade off risk and return. One potential way to resolve this problem is to adjust the charge for capital to consider the increased risk aversion of the agent relative to the agent. See Christensen et al. (2000) and Dutta and Reichelstein (2000) for analyses where the agent is risk averse, and investment levels affect potentially both market-related risk (which both the principal and the agent must bear) and project-specific risk (which only the manager bears. These papers find that the principal must make the charge for capital lower than his own risk-adjusted cost of capital in order to give the agent incentive to bear project-specific risk. It is important to note that when there are incentive problems involving risk–return tradeoffs, the assumption that the compensation contract is linear becomes suspect. In particular, convex contracts can be used to offset the concavity of the agent's utility function and therefore induce the agent to behave in a less risk averse fashion. See Lambert (1986), Meth (1996), Feltham and Wu (2001), and Demski and Dye (1999) for models where the principal must motivate risk–return tradeoffs.

expected utility (assuming s is normally distributed) is

$$\frac{\mathrm{E}(s) - 0.5\rho\,\mathrm{var}(s)}{1+r} - C(a) + V(I),$$

where ρ is the agent's coefficient of risk aversion. Note that the agent's monetary utility is also discounted at the rate of r, since he is paid at the end of the period, while his nonpecuniary returns (C and V) occur at the beginning of the period. We could also model the nonpecuniary returns as being defined at the end of the period with no loss of generality. The agent's reservation level of expected utility is $\underline{H}$.

Finally, assume that the end-of-period cash flow is not necessarily available for contracting. Instead, the principal and agent jointly observe a "noisy" measure of the cash flow, $\tilde{Y} = \tilde{X} + \tilde{e}_y = ba + \mathrm{E}(x_I|I,m) + \tilde{e}_a + \tilde{e}_I + \tilde{e}_Y$. When the variance of e_y is zero, the model is the same as if the end-of-period cash flow is contractible.

In the absence of incentive problems, the principal's problem is to choose s, a, and I to

$$\text{maximize } \mathrm{E}\left(\frac{\tilde{X} - s}{1+r} - I\right) \tag{25}$$

$$\text{subject to } \frac{\mathrm{E}(s) - 0.5\rho\,\mathrm{var}(s)}{1+r} - C(a) + V(I) \geqslant \underline{H}. \tag{25a}$$

With a weakly risk averse agent and a risk neutral make principal, the principal will absorb all the risk, and the agent will be paid a constant. This constant will be the reservation utility constraint given by the equality $s/(1+r) = \underline{H} + C(a) - \mathrm{E}_m[V(I)]$. The optimal actions satisfy the equations

$$a = \frac{1}{1+r}b,$$

$$\frac{1}{1+r}\frac{\partial \mathrm{E}(\tilde{X})}{\partial I} + V'(I) - 1 = \frac{1}{1+r}\frac{\partial \mathrm{E}(\tilde{x}_I|I,m)}{\partial I} + V'(I) - 1 = 0.$$

Note that in the absence of incentive problems, the action choices are separable. That is, the optimal level of productive effort does not depend on the investment choice, and vice versa. The second equation implicitly gives the optimal investment level, I. Note that this investment will depend on the information signal m if this is informative about the marginal productivity of investment. Moreover, the agent's nonpecuniary return associated with the investment level is taken into consideration in selecting the optimal level of investment. For example, if the agent prefers higher levels of investment to lower levels, $V'(I) > 0$, the principal selects a higher level of investment because this allows him to lower the compensation needed to provide the agent with his reservation level of utility.

Now suppose the agent's actions are unobservable, and suppose he is compensated using a performance measure that can be written as

$$w = \tilde{Y} - \delta I.$$

This performance measure can be thought of as the estimated cash flow or net income (gross of the agent's compensation) if $\delta = 1$. That is, cash flow and income are the same in a single-period model since there are no depreciation issues.[64] Residual income can be represented using $\delta = 1 + r$. The agent's compensation is assumed to be a linear function of the performance measure:

$$s = \beta_0 + \beta_1 w.$$

Suppose the agent is given a contract (β_0 and β_1) and a charge for capital (δ). The agent's problem is to

$$\underset{a,I}{\text{maximize}} \frac{\mathrm{E}(s) - 0.5\rho \, \text{var}(s)}{1 + r} - C(a) + V(I),$$

which is

$$\frac{\beta_0 + \beta_1[\mathrm{E}(x_a|a) + \mathrm{E}(x_I|I,m) - \delta I] - 0.5\rho\beta_1^2 \, \text{var}(x_a + x_I + e_y)}{1 + r}$$

$$- C(a) + V(I).$$

The agent's first-order conditions are

$$a = \frac{\beta_1}{1 + r}b,$$

$$\frac{\beta_1}{1 + r}\left[\frac{\partial \mathrm{E}(x_I|I,m)}{\partial I} - \delta\right] + V'(I) = 0.$$

Note that the agent's optimal investment (I) depends on both the magnitude of the slope coefficient (β) in the compensation contract and the charge for capital, δ. The greater the charge for capital, δ, the less the agent will invest. Therefore, the agent invests less when the performance measure is residual income than when it is cash flow or net income, ceteris paribus.

As usual, the agent's acceptable utility constraint will be met with equality and we can substitute for $\mathrm{E}(s)$ into the objective function[65]

$$\text{maximize } \mathrm{E}\left(\frac{\tilde{X}}{1 + r}\right) - \underline{H} - C(a) - 0.5\rho\beta^2 \, \text{var}(\tilde{Y}) + V(I) - I \qquad (26)$$

---

[64] Depreciation issues will be discussed in a later section.

[65] When the agent's investment choice depends on the private information signal, m, the ex ante distribution of his wealth is no longer likely to be normally distributed. While this does not affect the incentive compatibility constraints, it does affect how the agent's acceptable utility constraint is modeled. I ignore this issue here because I do not believe it likely to alter the primary conclusion: that the incentive problem with the productive effort interacts with the incentive problem on investment such that the optimal charge for capital to the agent need not equal the principal's cost of capital.

subject to $a = \frac{\beta_1}{1+r}b,$ (26a)

$$\frac{\beta_1}{1+r}\left[\frac{\partial E(x_I|I,m)}{\partial I} - \delta\right] + V'(I) = 0.$$ (26b)

The solution to this problem is straightforward and has the following characteristics:

(i) The first-best level of short- term investment, I, is selected.
(ii) The optimal slope coefficient is $\beta_1 = [b/(1+r)]^2/[\rho \operatorname{var}(\tilde{Y}) + [b/(1+r)]^2].$
(iii) The optimal charge for capital is $\delta = (1+r)\big(1 + ((1-\beta_1)/\beta_1)V'(I_{fb})\big).$

This proposition shows that if the charge for capital is a choice variable, it can be chosen to motivate the agent to select the level of investment level that is best for the principal. This occurs even if the agent has a nonpecuniary return associated with the level of investment. The equation for the slope coefficient in the contract is exactly the same as it would be in a model with no investment (other than the effect of the increase in the variance due to the cash flow generated by the investment). Therefore, the coefficient β_1 is used solely to motivate the effort a. Note that if $\rho = 0$ or $\operatorname{var}(Y) = 0$, we have $\beta_1 = 1$. Otherwise $0 < \beta_1 < 1$. Given the slope coefficient in the contract, the charge for capital δ is then adjusted to motivate the correct level of investment. This occurs because there is no "risk effect" associated with the charge for capital (unlike the slope coefficient β_1). Therefore, the charge for capital can be adjusted to whatever the principal wants to get the desired level of investment, and this charge for capital has no other effects.

However, the optimal charge for capital must consider the incentive problem on productive effort (a). The form of this dependence is specified as follows:

(i) If the agent is risk neutral ($\rho = 0$) or if the agent has no nonpecuniary returns for the level of investment ($V'(I) = 0$), the optimal charge for capital is $\delta = 1 + r$.
(ii) If the agent is risk averse and $V'(I_{fb})$ is positive, the optimal capital charge is higher than $1 + r$.
(iii) If the agent is risk averse and $V'(I_{fb})$ is negative, the optimal capital charge is lower than $1 + r$.

Under the first set of conditions, evaluating performance using residual income yields the first-best level of investment. When the agent has no nonpecuniary return associated with the level of investment (as in the Rogerson (1997) and Reichelstein (1997, 2000) papers), the agent cares only about the financial impact of the level of investment. To get him to view the investment

choice in a way that is congruent with the principal's view, the charge for capital in the agent's performance measure must be the principal's cost of capital, $1 + r$. Note that this works regardless of the slope coefficient in the agent's contract (as long as it is not zero). Moreover, the principal need not know anything about the shape of $E(X_I | I, m)$ for this to work. Therefore, the principal does not need to know the information signal m or have the agent communicate this information.

Similarly, if the agent is risk neutral, it is optimal to "sell the firm" to the agent by setting the slope coefficient equal to 1. This forces the agent to internalize the production effort problem. To get the agent to make the correct investment choice, he must also internalize the cost of capital the principal bears. This is achieved by setting the charge for capital to the agent to equal the principal's cost of capital, $1 + r$.

In contrast, if $V'(I) \neq 0$ and the agent is risk averse, the optimal charge for capital is not the same as the principal's cost of capital. If the agent prefers higher levels of investment (for the prestige), the charge for capital must be greater than the principal's cost of capital. Since the agent receives all of the prestige, but only portion of the financial returns to the project, he has an incentive to overinvest. Therefore, the principal has to raise the charge for capital in the performance measure to prevent the agent from overinvesting. Similarly, the optimal capital charge is lower than $1 + r$ if $V'(I_{\text{fb}})$ is negative. That is, if the agent prefers lower levels of investment (so he does not have to work as hard), the principal has to lower the charge for capital to induce the agent to invest more.

Note that for general forms of $V(I)$, the principal must know the functional forms of $E(x_I | I, m)$ and $V(I)$ to correctly set the charge for capital that leads to first-best investment. However, for the special case where $V(I) = vI$, we have $V'(I) = v$, and the optimal charge for capital is

$$\delta = (1 + r)\left(1 + \frac{1 - \beta_1}{\beta_1} v\right).$$

Now the principal does not have to know the form of $E(x_{1S} | I, m)$ or be able to calculate I_{fb} to set the correct charge for capital. Therefore the principal does not need to know the information signal m or have the agent communicate it. However, even here the principal must consider the incentive problem on the agent's productive effort (which determines the magnitude of β_1) to decide what the appropriate capital charge should be.

Finally, note that the optimal charge for capital depends on the variance of the accounting earnings at the end of the year. That is, if the variance of the real cash flow increases *or* the variance of the noise in accounting numbers increases, the slope coefficient in the agent's compensation contract goes down. When the slope coefficient goes down, the charge for capital must adjust accordingly. Consider the case where $V'(I) > 0$, so the agent prefers higher

levels of investment, ceteris paribus. In such a setting, the charge for capital is higher than the principal's cost of capital. When the agent's slope coefficient is reduced, the principal must increase the charge for capital even more to control the agent's incentives to overinvest. Therefore, increased noise in earnings translates into a higher charge for capital. Similarly, when $V'(I) < 0$, the opposite result occurs. That is, increased noise causes the principal to decrease the charge for capital (which is below the principal's cost of capital to begin with) in order to motivate the agent to select the correct level of investment.

The link between accounting systems and the internal cost of capital is a fruitful one for future research. Clearly, the role of "managerial" accounting systems and disclosures in helping or hindering the problem of capital allocation inside the firm is conceptually similar to the role of the "financial" accounting and disclosures in helping investors to allocate capital across firms. Therefore, this work should also have implications for the external cost of capital for firms.

4.4. Transfer pricing

There are many parallels between the agency literatures on capital budgeting and transfer pricing. In part, this is because both are a form of resource allocation within firms. In capital budgeting the resource is capital, and it is transferred from headquarters to divisions. In transfer pricing, the resource is an intermediate good that is transferred between an "upstream" division to a "downstream" division. In fact, the capital investment models of Antle and Eppen (1985) and Antle and Fellingham (1997) are outgrowths of Harris et al. (1982) classic paper on transfer pricing.

HKR model an organization which consists of a principal, one upstream division, and N downstream divisions. The upstream divisional manager can use capital supplied by the principal and his own effort to produce an intermediate product. The downstream divisions use the intermediate product and their own effort to produce products sold to consumers. Each divisional manager has private information about the relative productivities of effort versus the resource supplied by the previous stage in the production process. Moreover, each manager has an incentive to overstate the benefits of the resource supplied by the prior stage in order to reduce the amount of effort he has to supply.[66]

HKR's results are qualitatively similar to those in Antle and Fellingham (1997) in that both "rationing" and "slack" are part of the optimal solution. The transfer price to downstream agents is set to be higher than the cost of the

---

[66] While the modeling is different, this is similar in spirit to the assumption that manager's derive utility from obtaining more resources than they really need to perform their job, i.e., they can consume "slack".

resource in order to reduce the downstream agents' incentives to request more resources than they actually need. This is in sharp contrast to the prescriptions from the neo-classical literature on transfer pricing, which suggest that resources be transferred at their marginal cost.

Setting transfer prices at marginal cost is also unlikely to be optimal in situations where agents can make "relation specific" investments which lower the marginal cost of production (or in the case of an end-unit division, raise the marginal revenue or net realizable value of the product). When divisions must share the benefits of these investments but bear the full costs because they are noncontractable (either because they are nonpecuniary investments or are monetary investments which cannot be observed because they are difficult to untangle from other investments made by the division), they will underinvest.

The transfer pricing literature has examined a number of different incentive problems: the incentive for divisions to make upfront investments that improve their marginal profitability (Alles and Datar, 1998; Sahay, 1997; Sansing, 1999; Baldenius et al., 1999; Baldenius, 2000), conflicts over the optimal production quantity, tradeoffs between using the resource transferred in versus the agent's own effort (HKR), allocation of effort between products that are sold externally versus products that are transferred internally, and tax effects of transfer pricing versus incentive effects (see Smith, 1999).

Many of the more recent papers in transfer pricing have moved away from deriving the optimal transfer pricing mechanism, and they have instead focused on comparing specific alternative transfer pricing (e.g., cost-based transfer pricing versus negotiated transfer pricing systems). Unfortunately, it is difficult to compare the results across studies because they frequently use different definitions of cost-based or negotiated transfer prices. For example, in Baldenius et al. (1999) the negotiation determines how the two divisions split the realized overall contribution margin, whereas in Baldenius (2000) the two divisional managers negotiate an ex ante transfer price. As this work progresses, a greater uniformity in models and definitions will facilitate comparisons.

Future work in transfer pricing should keep in mind that the contracting relationship between "independent" firms differs in very significant ways between the contracting problem within firms.[67] First the transfer price between divisions is not a real transfer of money; it is generally merely an accounting charge on the divisional books. By the same token, the divisional manager does not have property rights over his divisional income; the profits of the division are not his to consume. The only reason he cares about his divisional profits at all is if the principal chooses to evaluate him and

---

[67] See Baiman and Rajan (2000) for a recent review of contracting relations between firms, and Demski and Sappington (1993), Baiman and Rajan (2001), and Kulp (2000) for recent examples of papers.

compensate him on the basis of this performance measure.[68] In some papers, it is not obvious why the principal would choose to do this. In particular, if the agents were evaluated and compensated on the basis of firm-wide profits, the incentive problems modeled would disappear.

Future work should also more fully explore the use of dual prices (i.e., transfer pricing systems where the price credited to the upstream division need not be the same as the price charged to the downstream division). Dual pricing would seem to eliminate the problems induced by the "zero sum" nature of conventional transfer pricing mechanisms. For example, under dual pricing, giving an upstream division a high transfer price in order to motivate the manager to make relation-specific upfront investments would not cause the manager of the downstream division to decrease the quantity of the intermediate good he demands. Identifying the costs of a dual pricing scheme (does it lead to bigger incentives for managers to collude?) would be an interesting area for future research.

4.5. Cost allocation

There are also many similarities between transfer pricing and cost allocation; in fact, cost allocation can be viewed as a special case of transfer pricing. Nevertheless, there are differences in how the two literatures have progressed. The early agency literature emphasizes that the value of a cost allocation system comes from its ability to construct a more informative performance measure about the agent's actions (see Demski, 1981; Magee, 1988). However, while transfer pricing models seem to focus on a single downstream division, the cost allocation literature seems more likely to consider multiple users. In a multiple user setting, Rajan (1992) shows that cost allocation schemes can also be valuable in reducing the users' ability to collude to misreport the productivity of the resource to them.

Moreover, relatively few transfer pricing models are set up in a way that allows for a separation of production costs into a fixed and variable component. This precludes these studies from examining full-cost-based transfer pricing policies, which many surveys claim to be widely used. In contrast, a number of cost allocation papers investigate the optimality of allocating full costs.[69] An important issue in analyzing the effects of full cost allocation is the dynamic nature of the resource allocation acquisition and use. The allocation system must motivate agents to acquire the proper amount of productive capacity in the first place (this may involve having agents forecast

---

[68] While divisional profits are undoubtedly one input into the manager's compensation function, many divisional managers are also compensated on the basis of segmental or firm-wide profits and also on the basis of nonfinancial measures such as meeting divisional objectives.

[69] see Miller and Buckman (1987) and Hansen and Magee (1993).

their expected usage) and to efficiently allocate the resource once it has been acquired. It is not clear that a cost allocation system can do both (see Hansen and Magee (1993) for discussion). Problems in allocating the "fixed" portion to users include (i) a large component of the fixed cost is a sunk cost by the time it is used, (ii) the opportunity cost of using the resource depends on how many others desire to use it; costs of congestion are highly nonlinear, (iii) the acquisition of capital intensive production resources is frequently "lumpy", and (iv) new information about the profitability of using the resource to the users may have arrived since the time the resource was originally acquired.

Relatively little attention has been placed on who (if anyone) pays for "idle" capacity or for "excess" inventory at intermediate stages of the supply chain, or how penalties for underacquisition of a resource or underproduction of an intermediate good are levied. These represent fruitful areas for future research.

4.6. How does the agent acquire private information?

In the previous sections, the models implicitly assumed the agent acquires the private information "automatically" as part of the job. We now discuss situations where the principal or the agent can influence the type of information produced. For example, the principal could choose what kind of decision-support system to install for the agent's use, including the variables to provide the agent, as well as the level of detail provided. Alternatively, the agent could spend some of his effort collecting and processing information before he decides how to make use of the information. Relatively little research has addressed these kinds of problems, and the results we have are hard to relate to each other.

At present, we do not have a very good understanding of when the principal is better off providing the agent with a system that generates private information. Christensen (1981) provides an example where the principal is better off and one where he is worse off when the agent receives private information relative to the case where both parties remain "uninformed". When the principal is worse off, it is because the agent is able to use the information to know how much slack he can get away with.[70] Baiman and Sivaramakrishnan (1991) provide additional examples of situations where the

---

[70] Specifically, suppose $x = m + a + \varepsilon$, where $m > 0$ is the private information observed by the agent, $a \geq 0$ is the agent's effort, and $\varepsilon \geq 0$ is the residual uncertainty. In the first-best solution, the optimal effort does not depend on the information signal m. If neither party can observe the signal m, the first-best solution can still be achieved using a forcing contract because the lower bound on the outcome is the agent's action, a. That is, the outcome distribution exhibits moving support. The forcing contract penalizes any outcome below $x = a$. With a big enough penalty you can motivate the agent to pick the first-best action and never have to pay the penalty. However, if the agent can observe the signal before he selects his effort level, then for all signals $m > 0$ the agent can reduce his effort level by m and not get caught. Therefore, the first-best solution cannot be obtained.

principal is worse off even though the agent's effort levels are (weakly) higher under the private information regime.

In contrast, Christensen (1981) and Penno (1984) formulate examples where the principal is better off giving the agent access to private information because the agent's information tells him about the marginal productivity of his actions. When he finds out effort is not productive, he provides none. But when he finds out effort is productive, he provides more than he would have in the no information (average productivity) case. This allows the principal to tailor the strength of the incentives and the risk being imposed to the productivity of effort. However, we are not close to having very broad sufficient conditions for information to have value.[71]

When the firm's outcome is not available for contracting purposes (so that an imperfect performance measure must be used in its place), determining the value of providing the agent with pre-decision information is more complicated. Bushman et al. (2000) (BIP) show that the value depends on both the correlation structure between the agent's private information and the real outcome and the correlation structure between the agent's private information and the performance measure used in the contract. A strong correlation on any one link is not sufficient for the information to have value. If only the first correlation exists, the manager will ignore the information because it does not affect his compensation. If only the second correlation exists, the manager will use the information signal to "game" his effort decision at the principal's expense.

Given that both links exist, do the principal's expected profits improve by more if he provides the agent with a pre-decision information system that is more accurate on the first link or the second? BIP show that the choice can go either way; in particular, they provide examples where the principal is better off by improving the link between the manager's information and the observed performance measure than by improving the manager's information about how his actions will affect the real cash flow. Their results also suggest that a private information system that results in an equilibrium where there is a high *unconditional* correlation between the real outcome and the performance measure is not as valuable as one that generates a high correlation *conditional* upon the manager's private information signal.

An alternative way to model the information acquisition process is to assume that the agent has to work to collect information, process it, and determine what the consequences of different courses of actions will be. Lambert (1986), Demski and Sappington (1987), examine the interaction between motivating the agent to work to collect information and motivating him to then use this information to make good decision on the principal's behalf. Lambert shows

---

[71] See Baiman and Evans (1983) and Baiman and Sivaramakrishnan (1991) for additional conditions where giving the agent private information makes the principal better or worse off.

that the principal must make the agent's compensation depend on the firm's outcome to give the agent incentive to work to collect information, but this risk affects the agent's incentives to adopt projects that are risky. He develops conditions where the optimal contract motivates the agent to be overly conservative in his risk–return choices and conditions where the agent is motivated to be overly risk seeking.

4.7. Value of communication

In models where the revelation principle applies, it is always weakly valuable to allow the agent to communicate his information to the principal. However, we do not have much knowledge regarding the factors that cause communication to have strictly positive value or the factors that affect the magnitude of the value of communication. In some cases, we know that the value is nonmonotonic in the precision of the information. At the one extreme, if the agent has no private information at all (the precision is zero), there is no value to communication. At the other extreme, Melumad and Reichelstein (1989) have shown that the value can also be zero when the agent has perfect pre-decision information (i.e., the agent faces no residual uncertainty). In their model the principal is able to infer from the outcome what the agent's private information was, so communication becomes redundant. For intermediate levels of precision, communication has positive value. At present, we do not have much knowledge about the shape of the value of communication as a function of its precision (or other parameters).

In multi-period models, we also have little information about the value of the *timing* of communication. In particular, how do we trade off timeliness versus accuracy of the information? Obviously these comparisons must address the question of what the communicated information will be used for. In most models analyzed to date, the only use is in contracting with the agent. However, it is likely that this information will also be valuable for other purposes. Even though one of the earliest applications of agency theory to accounting was on this issue (e.g., Gjesdal, 1981), relatively little work has examined the value of accounting information in alternative uses. See Narayanan and Davila (1998) for recent analysis on the tradeoffs between the cost and benefits of using a performance measure for compensation purposes versus other uses. In their model, the tradeoff arises because using the performance measure for compensation purposes motivates the manager to distort the information reported, which lessens its value for other purposes.

5. Earnings management versus the revelation principle

Earnings management is viewed as an activity that is widely practiced by managers. Even though the agency framework seems to be a natural one to use

to study earnings management, the agency literature to date has not made much progress in helping us understand how, why, and when earnings management takes place. The primary obstacle has been the revelation principle. As discussed in the previous section, when the revelation principle applies, any equilibrium that involves nontruthful reporting (i.e., ones where earnings management is taking place) can always be weakly dominated by one where truth telling is induced. It is only recently that researchers interested in nontruthful reporting have begun to construct models using features that ensure the revelation principle does not apply. While this does not guarantee nontruthful equilibria will be optimal, it at least opens the possibility.

There have been three different ways researchers have incorporated features designed to circumvent the revelation principle.[72] The most straightforward way is to simply exogenously restrict the agent's ability to communicate his information. Alternatively, some models place restrictions on the principal's ability to use the information, e.g., by requiring the principal to use a contract with a pre-specified shape (e.g., piece-wise linear). Finally, researchers have relaxed the assumption that there is pre-commitment as to how the agent's report will be used.

5.1. Communication restriction and costs

Communication restrictions and costs can relate either to limits on the manager's ability to misreport or limits on the ability of the manager to truthfully communicate his information. When either type of communication restriction exists, it is possible for nontruthful reporting to exist as equilibrium behavior.

Restrictions of the first type will often lead to higher expected profits than if these restrictions did not exist. To see this, suppose the agent privately observes the outcome at the end of the period, but he has unlimited discretion in what outcome he reports. If the agent's report is the only variable available for contracting, the report is useless for compensation purposes. Regardless of what the true outcome is, the agent will always report the same compensation-maximizing report. For example, if the compensation function is increasing in the reported outcome, the agent always reports the highest possible level of reported outcome. Moreover, since the probability distributions of the agent's reports and the compensation he will receive do not depend on his effort, he provides the minimum possible level of effort. The principal will anticipate this behavior, and he will offer the agent a flat compensation function (one where the agent's compensation does not depend on his report). Under this compensation plan, the agent is willing to report the outcome truthfully. Note that the communication problem and the incentive problem on the agent's

---

[72] See Arya et al. (1998) for additional discussion.

actions are intertwined. The principal is able to motivate truthful reporting, but at the expense of not being able to provide any incentives for the agent to work hard.

Now suppose there is a limit as to how much misreporting the agent can do. In particular, suppose that if the true outcome is x, the agent can issue a report in the range $[x - c, x + c]$ without his misreporting being detected. In this case, the agent will misreport earnings, but the principal will be able to invert his report to infer what the real outcome is. For example, if the contract is increasing in the reported outcome, the agent will always over-report the outcome by the maximum amount, but the principal knows that a report of m corresponds to a real outcome of $m - c$. In fact, the principal can adjust the parameters of the contract to replicate the expected profits he would receive if the agent's report was somehow constrained to be truthful. Unlike the previous case, the principal is able to motivate a positive level of effort, but the equilibrium does not involve truth telling.

Evans and Sridhar (1996) generalize this example by modeling the amount of reporting discretion as a random variable that only the agent knows.[73] As in the previous case, it is generally optimal to allow the agent some ability to manipulate earnings because it is too expensive to motivate truth telling for every possible (real outcome, amount of discretion) pair. However, when the amount of reporting discretion is a random variable, the principal cannot unambiguously infer the real outcome from the reported outcome. As a result, there is a welfare loss relative to the case where truthful reporting was exogenously imposed. Evans and Sridhar (1996) also analyze the same model in a multi-period setting, and they show that the "adding up" constraint of accrual accounting helps the principal control the agent's reporting behavior. That is, if the agent over-reports (under-reports) the outcome in the first period, he must under-report (over-report) the outcome by the same amount in the second period. This limits, but does not completely eliminate, the agent's incentives to distort the reported income.

In the examples above, the amount of reporting discretion is exogenously given. It is also possible to model the amount of discretion as endogenously determined. For example, as in the discussion of "window dressing" in Section 3.3.1, suppose that after the manager observes the real outcome x, he can exert "manipulation effort" to alter the reported outcome. In particular, suppose an effort level of c will increase the reported outcome to $x + c$, but the agent incurs disutility of $V(c)$ to exert this level of manipulation effort. The agent will choose a nonzero level of manipulation for at least some values of x, and the amount of his manipulation will be an increasing function of the slope

---

[73] This can also be considered a model of restrictions on the ability of the agent to communicate the truth. That is, the agent observes two random variables: the true outcome and the amount of reporting discretion. However, he is only permitted to communicate a one-dimensional signal.

coefficient of the compensation contract near the proposed level of reported earnings.[74] However, even if the principal can undo the amount of manipulation (e.g., if the value of c is independent of x), there is a welfare loss relative to the situation where truth telling is exogenously imposed. The welfare loss arises because real resources are used in manipulating earnings. The principal either pays for these directly or pays indirectly by having to provide the agent a higher level of compensation to meet his reservation utility constraint.

In the example above, the *direct* communication cost only exists if the manager chooses to manipulate the reported earnings number; there is no direct communication cost if the manager tells the truth. Not surprisingly, earnings management also occurs if there are direct costs to reporting the "truth". For example, suppose the "unmanaged" earnings number contains noise that the manager can observe, but that it is costly to take actions to remove this noise. The principal can only observe final report, but not the original earnings number or whether the manager has intervened to remove the noise. Verrecchia (1986) shows that it is optimal to let the manager decide when to remove this noise. In essence, it is too costly for the principal to induce the manager to always incur the cost of removing the noise.

A second type of communication restriction is that the agent cannot fully communicate the full dimensionality of his information. Managers often observe an extremely rich information set that would be very difficult and costly to communicate. Moreover, the principal will generally not have the technical expertise to be able to understand many of the dimensions of the agent's information set. When these types of restrictions exist on the manager's ability to communicate his information, the revelation principle fails virtually by definition. Now, the reporting problem contains an aggregation dimension as well as a misreporting dimension.

To illustrate, suppose that at the end of the period, the agent privately observes the cash flow for the period and an information signal related to the profitability of future period actions. If the agent's report can be only one dimensional (e.g., he reports accounting earnings), how would the principal like the agent to aggregate the two signals into one report? In some instances, the principal might prefer the agent to report on only one of the dimensions; in this case the distinction between truthful versus nontruthful reporting is clear. However, it is likely that the principal will prefer a report that combines both dimensions.[75] For example, the ideal reporting strategy could be to make reported earnings higher (lower) than the cash flow for the period when the

---

[74] See Fischer and Verrecchia (2000) for further analysis of the introduction of reporting bias by the manager.

[75] Gjesdal (1981) shows that, in general, the principal will not rank information systems identically for stewardship purposes as for investment purposes.

information signal about future period profitability is favorable (unfavorable). At present we have little knowledge of what aggregation rule the principal would like the agent to use or what incentives the agent would have to deviate from this aggregation rule.

5.2. Restricted contract form

Earnings management can also occur when the researcher exogenously restricts the form of the compensation contract. For example, Demski and Dye (1999) analyze a single-period model in which the agent is responsible for actions that affect the mean and variance of the end-of-period cash flows. At the beginning of the period, the agent observes private information about the mean and variance, and he communicates a report about both of these parameters to the principal. However, the contract is restricted to be a linear function of the end-of-period cash flow and a penalty term that is proportional to the square of the deviation of the realized cash flow from the forecasted mean.[76] They show that the manager's report always underestimates the expected cash flow. That is, in their model, the manager builds slack into his forecast. In contrast, in the reporting models that use optimal contracts such as Antle and Fellingham (1997) or Kirby et al. (1991), the manager's forecast is unbiased, and it is the principal who provides the slack to the agent as an inducement for the agent to forecast honestly.

In multi-period models, earnings management across periods will be affected by the shape of the compensation contract within a period and the structure of the compensation contracts across periods. Ceteris paribus, the agent's reporting behavior is driven by his incentive to equalize the marginal utility of consumption in the current period with his expected marginal utility in future periods. In many papers, the agent is assumed to not have access to capital markets to directly shift his consumption. Therefore, the only way for the agent to shift his consumption over time is to shift his compensation by "managing" the reported income of the firm across periods.[77] For the agent, lowering the reported income is equivalent to deferring some of his compensation.

---

[76] Demski and Dye model the manager as obtaining information about the mean and variance of multiple projects. Therefore, their contract is also based on the realized standard deviation of cash flows across projects and the deviation of this standard deviation from the manager's forecasted standard deviation.

[77] When the agent has the ability to borrow or save, we must compare these terms with the terms implicit in the compensation function. For example, agents generally will be able to save compensation, but they may not be able to borrow on very favorable terms. In deciding whether to loan money to agents, banks and other lending institutions face the same kind of moral hazard and adverse selection problems that principals face. If agents can save money, their incentives to decrease reported income go down.

To illustrate this, suppose the agent is risk averse and his utility function is additively separable over time. Further assume that his compensation depends only on the reported profits of each period and that the "unmanaged" profits are independently distributed over time. In this case, if the agent's compensation is linear or concave, he will want to adjust first-period outcomes that are extremely high (low) downward (upward). This behavior corresponds to "income smoothing". However, if his compensation is sufficiently convex, he will want to do the opposite: when the first-period outcome is very low, he will want to further decrease it because the extra compensation he receives from increasing a low outcome is small compared to being able to increase the reported outcome next period, when the outcome is likely to be higher. This behavior corresponds to "anti-smoothing".

For other shapes of compensation functions and other stochastic processes for earnings, more sophisticated strategies for managing earnings could arise. For example, in Healy's (1985) analysis of earnings management, the compensation scheme is assumed to be piece-wise linear to yield an "S" shape. Healy also implicitly assumes that the terms of the contract in the second period are not adjusted based on the first-period outcome. This structure leads to the prediction that the agent will decrease reported earnings if earnings are extremely high or low (i.e., when he is on the flat part of the compensation function), and he will increase intermediate levels of earnings (when he is on the sloped part of the compensation function).

While exogenously specifying the form of the contract can be helpful in understanding the reporting incentives implied by that structure, it begs the question of why the compensation is structured that way in the first place. At present, we have very little knowledge for why compensation contracts have the shape they do: why components are piece-wise linear, or why bonuses are "lumpy, i.e., a full bonus is paid if performance exceeds a threshold, and no bonus at all is paid even if performance barely misses this threshold. Casual empiricism suggests that contracts with these features are very common, and that these features have a major effect on managers' incentives. The linear contracting framework is incapable of providing insights into these issues; a more general (ideally, an optimal) contracting framework is necessary. We also have very little knowledge about how the terms of contracting evolve over time (and exception is Indjejikian and Nanda's (1999) model in which the second-period contract employs the "ratchet" effect").

5.3. Inability to precommit to how the information is used

The pre-commitment assumption is important to the revelation principle because the principal's promise to "under-utilize" the information is what gives the agent the incentive to reveal the truth. If the agent believes the information will be used against him, it becomes more costly (perhaps too costly) to

motivate him to reveal the truth. One way in which the pre-commitment assumption has been relaxed is by assuming there are other parties who see the agent's report who cannot precommit to how they will use it. For example, this third party could be an auditor who is hired to issue an opinion about the agent and his reports. Baiman et al. (1987) analyze a model in which they show it cannot be an equilibrium for the agent to always tell the truth and for the auditor to work hard to investigate the agent's report. That is, if the auditor is convinced the agent's report is truthful, he has no incentive to spend time and money doing the audit. In theory, the third party could also be another employee of the firm, a competitor, the labor market (in setting the value of the agent's outside employment opportunities in the future) or even the stock market (in setting the value of the agent's stock-based compensation), and the revelation principle would not apply.

A second way to violate the pre-commitment assumption is to assume the principal himself cannot precommit to how he will use the agent's report. Generally speaking, models of this type are multi-periods ones. While the principal is able to pre-commit to how the agent's report will be used in his first-period compensation, he is not able to make a similar commitment regarding how the first-period report will affect the agent's second-period compensation.[78] One interpretation to this is that the principal offers a two-period contract that he is later able to renege on if it is in his best interests to do so. An alternative interpretation is that the principal can only offer one-period contracts, that employment contracts are "at will" (see Arya et al., 1997). Demski and Frimor (1999) show that the revelation principle can be avoided even if the principal can write a credible two-period contract, as long as the principal and agent can renegotiate its terms in the second period. In these models, the agent anticipates that the principal will later use his report against him, and this makes it more expensive to report the truth.

In fact, in many of these models, the threat of renegotiation is so severe that the principal and agent are better off ex ante if they set up a system that prevents the agent from issuing *any* report at the end of the first period.[79] This is in striking contrast to single-period models or models where multi-period pre-commitment strategies are credible. In particular, communication need not have positive value, information delay can make both parties better off, and aggregation of information can actually improve both parties'

---

[78] In these models, it often seems arbitrary as to what things the principal is allowed to precommit to and what things he is presumed unable to precommit to.

[79] While the reasons are somewhat different in different papers, they are generally related to the idea that the release of information decreases the ability of the principal and the agent to insure the agent against risk. That is, risks that are ex ante optimal to insure against become ex post optimal to renegotiate.

welfare.[80] Therefore, these papers have taken us from one extreme (truthful reporting is always weakly optimal) to the other extreme (issuing no reporting is optimal). It would be interesting and more descriptive to explore models which yield a nontrivial, yet not completely truthful, first-period reporting strategy. To do this, there has to be a more substantive role for "early reporting" to play. I discuss these and other multi-periods issues in the next section.

6. Multi-period models and investment problems

While there are a number of interesting issues that arise in multi-periods agency models, the one I believe is of greatest interest to accounting relates to the role of lead–lag effects in performance measures. For example, we need a multi-period model to be able to talk about accrual accounting problems because in single-period models, cash flow and accrual accounting numbers are identical. Despite the obvious importance, not much work has been done on multi-period models in the agency literature. The reason is tractability problems. In most multi-period models, numerous technical issues arise that are often tangential to the accounting or performance measurement issues that we would like to focus on as accounting researchers. For example, even with models where everything seems to be independent over time, we have to worry about borrowing and lending, wealth effects, randomization, how the contact parameters in one period depend on realizations from prior periods, the form of the contract, the ability to commit to long-term contracts, etc.[81] Even information signals that would seem to be informationally "meaningless" sometimes play an important role in helping to randomize actions or in coordinating the actions of different parties.

As discussed previously, some papers have addressed multi-period issues in an ad hoc fashion by analyzing models that are really one-period models, but in which the true outcome is not observed till beyond the contracting horizon. For example, in these models, some of the agent's actions have only a "short term" effect that is properly captured by that period's accounting earnings, whereas other actions have longer term effects that the current period accounting number does not capture. It is encouraging to see that a number of recent papers have made progress in being able to formulate genuine multi-periods models tractable enough to solve and interesting enough to yield insights.

In addition to the multi-periods models of communication where the revelation principle breaks down as discussed in the previous section, the other

---

[80] See Demski and Frimor (1999), Indjejikan and Nanda (1999), Arya et al. (1997), and Gigler and Hemmer (1998).

[81] See Lambert (1993), Rogerson (1985b), Fellingham et al. (1985), and Fudenberg and Tirole (1990) for analysis of some of these issues.

main branch of the multi-period literature focuses on motivating long-term investments. These papers compare alternative financial measures of performance such as operating cash flow, accrual accounting net income, residual income, and discounted expected future cash flow. Of particular interest is Rogerson (1997) and Reichelstein (1997, 2000), and Dutta and Reichelstein (1999). Their results lend theoretical support for the use of residual income or economic value added (EVA) as a measure of performance. In fact, the results in these papers are strikingly strong.

In these papers, the agent is responsible in each period for selecting an investment level, I_t, as well as perhaps selecting a "productive" effort level, a_t. The investment yields returns that are spread out over the project life, while the productive effort yields an immediate return. The agent has private information, m_t, about the marginal profitability of the investment. The primary result in these papers is that the agent can be motivated to select the "first-best" level of investment if his compensation is based on the firm's residual income. The charge for capital to be used in calculating residual income is the principal's cost of capital. The charge for capital is applied to the investment's book value, which is the original investment minus any accumulated accounting depreciation. The papers shows how to "correctly" measure the book value of the investment (i.e., how to calculate the "correct" depreciation schedule).

Surprisingly, motivating the agent's investment does not require any restrictions on the form of the compensation function (only the performance measure used). Therefore, the principal is free to adjust the form of the compensation scheme to deal with other kinds of incentive problems he faces (e.g., motivating the agent to work hard). Moreover, this result holds (i) regardless of the agent's time preferences (the agent can have a shorter time horizon than the principal), (ii) regardless of the agent's utility function, and (iii) without the principal observing m or having the agent communicate any information about the realization of m.

These results are so strong relative to other agency theory results that they seem too good to be true. As I argued in a previous section, in part these results occur because the agents are assumed to care only about the monetary effect of the investment; they receive no nonpecuniary returns (or bear no nonpecuniary costs) associated with the level of investment.[82] When this assumption is relaxed, the first-best level of investment can be achieved, but the correct charge for invested capital is generally not the principal's cost of capital, and the correct cost of capital must be adjusted in response to the other incentive problems present in the model.

These papers also make strong assumptions about the information available about the time pattern of the returns from investment. When an investment

---

[82] Another reason is because there are no risk–return tradeoffs the agent must make.

generates returns that are spread out over many periods, accounting income and residual income measures must determine a way to depreciate the investment over time. In order to get residual income to correctly motivate the agent's investment choice, the principal must calculate the "correct" depreciation schedule. To do this, the principal must be able to "match" the depreciation to the time pattern of the cash flows generated by the investment.[83] In the investment papers referenced above, the principal is able to do this because he has complete information about the time pattern of the cash flows. The agent's investment decision and his private information signal influence the scale of returns, but not their time pattern.[84] The realizations of the cash flows in each period therefore provide no new information about the profitability of the investments, and the principal only needs to know how large the agent's investment choice was to be able to calculate the correct depreciation schedule.

While this is an interesting and important benchmark case, clearly it would be useful for future work to relax this assumption. In particular, we could allow the agent to have superior information about the profitability or time pattern of returns or to be able to take actions to influence the principal's perceptions of these things. For example, the agent could keep the reported overall profitability constant but understate expected cash flows in early years and overstate them in later years. By understating them in early years, he could receive a "favorable" depreciation schedule as well as look good by beating the forecast in the early years. His shorter time horizon allows him to avoid paying the cost when this gets undone in the later years.

7. Directions for future research

I have attempted to indicate unresolved or unexplored issues at various places in the paper; however, in this section I will outline what I feel are the most important areas and issues for future research to explore. The first area relates to the aggregation of performance measures. A fundamental property of accounting systems is that they aggregate "basic" signals. Moreover, the aggregation is done in specific ways; in most cases, the aggregation is linear and all dollar amounts are weighted equally (other than revenues and expenses

---

[83] This is similar to the calculation of "economic" depreciation.

[84] For example, Rogerson (1997) assumes the return in period t from an investment of I after observing an information signal m is $x_t = z_t R(I, m) + \varepsilon_t$, where the time patterns of z_t are known to both principal and agent. In Dutta and Reichelstein (1999) the present value of the return from the agent's investment (in sales effort) can be calculated in the period of the sale. The realized series of cash flows is a zero net present value series added to this present value. In both cases, the realization of the cash flow in one period does not provide any information about the future cash flows.

being weighted with opposite signs). Are these features consistent with optimal performance measurement? If so, why? In single-person decision theory, the decision maker is always (weakly) better off using the unaggregrated signals than using the aggregated ones. However, in some cases, aggregation does no harm if it is done correctly. The work of Banker and Datar (1979) is in this vein; they derive conditions where the optimal contract can be interpreted as aggregating signals in a linear fashion. However, unless severe restrictions are placed on the model, signals are not weighted equally. Perhaps if the model was expanded (instead of restricted), we might find that equal weighting of signals is a robust result. That is, if the agent's action space was expanded to allow him to take actions that increase one measure at the expense of another, these arbitrage opportunities might force the principal to weight the signals equally.

An alternative route is to explore models where aggregation has real benefits. One possibility is to consider the costs of processing a large number of disaggregated signals. These costs could either be human information processing costs (something economics-based research in general and agency theory in particular has had little success or apparent interest in doing) or administrative or legal contracting costs. See Dye (1985) for an example of introducing contract complexity costs into the analysis. Another relatively unexplored contracting cost is the cost imposed by renegotiating contracts or the inability to commit to long-term contracts in dynamic agency models. Indjejikian and Nanda (1999) show that aggregating information before it gets reported to the principal can reduce some of these costs. It would be interesting to explore these dynamic issues.

A second major issue for future research is understanding the process accounting systems use in transforming streams of cash flows into earnings numbers. The accrual process that makes earnings different from cash flow is a fundamental characteristic of accounting systems. To date, we have very few models which are rich enough to be able to incorporate features that allow us to address questions involving the matching principle, the choice between capitalizing versus expensing, the policy of being conservative or liberal, etc.

Related to the accrual accounting issue, a third major issue for additional work is earnings management. In order to be able to address the richness of earnings management strategies alleged to exist in practice, we need to move away from models in which the revelation principle applies. A subtopic in this area relates to the optimal shape of the contract. The shape of the contract (is it concave or convex in a given region, does it contain "jumps", etc.) will affect the agent's optimal reporting strategy. If the principal anticipates the agent's reporting behavior, why are contracts designed with the observed shapes?

Another relatively unexplored area for research is where the agent's (possibly manipulated) report is used by the market for valuation purposes as well as by the principal for compensation purposes. For example, it would be interesting to compare the market's ability to "undo" manipulations versus the

compensation contract's ability to do so. This may lead us to re-examine the role of stock-based compensation in motivating the agent's actions and his reporting behavior. Future work should also compare the roles of private information and communication in agency models versus models of disclosure in financial markets (as reviewed by Verrecchia, 2000). In agency models, disclosure is "mandated" by the principal but can be nontruthful, whereas in disclosure models the disclosure is typically "voluntary" but all disclosures are exogenously guaranteed to be truthful.

The topic of stewardship versus valuation uses of accounting information is also a good one for future research. While they are not identical, I suspect there is a closer relationship between these two things than our current models imply. For example, in our current models, constructing congruent measures of performances need not have any relation to constructing measures that are correlated with stock price or with underlying long-term value. In more dynamic models, where the productivity of actions varies over time, congruity and correlation may be more highly related concepts.

Another important but relatively unexplored area is the optimal charge for resources used by people within organizations. Capital budgeting, residual income calculation, transfer pricing, and cost allocation are subtopics within this area. In particular, the optimal charge for resources used for performance evaluation and incentive purposes may differ from the "correct" charge from a valuation perspective. Therefore, in contrast to my conjecture in the previous paragraph, this is a force that will cause stewardship and valuation measures to diverge. This makes identifying what the sources of divergence are and when they are significant all the more important for future researchers.

A major challenge confronting agency theory researchers in addressing many of these issues is constructing multi-period models which are tractable enough to solve but have interesting information and performance measurement issues. To do this, we need to make sure there is a reason to produce interim performance information and performance measures. That is, we cannot be able to solve the incentive problem (whether it is an investment one or a reporting one) by simply waiting till the end of the last period of the firm and paying the agent then.[85] Similarly, it should not be the case that we can simply pay the agent based on the realized cash flows each period and count on him being around until the end to resolve the incentive problem in this way. For example, it is often reasonable to assume that the agent has a shorter time horizon than the principal. This can be done by incorporating either a probability of the agent leaving (either voluntarily or exogenously) or a higher

---

[85] The research papers on residual income generally do not have this problem. However, a number of the communication papers do. In particular, the assumption of negative exponential utility over the sum of the compensation received means that there is no cost to waiting till the last period (when all uncertainty has been resolved) to pay the agent.

discount rate for money for the agent. When the agent has a shorter time horizon than the principal, there are benefits to having a "forward looking" performance measure. For example, if an investment project generates negative cash flows in the early years and positive cash flows in later years, he will be unwilling to invest the optimal amount if he is evaluated based on the cash flows that are realized during his tenure. There are a number of ways to calculate a forward looking performance measure, including deferring some of the investment's cost, by recognizing the future benefits, or by supplementing financial measures with nonfinancial measures, etc.

In many important multi-period problems, it will also be necessary to incorporate private information into the model. That is, the agent knows more about future period cash flow effects than the principal. How much discretion will the principal build into the accounting system to allow the agent's reports to depend on this private information? The shorter time horizon for the agent and the superior information of the agent set up the classic tradeoff in accounting: we would like accounting reports to reflect information that is forward looking, but forward-looking information is less reliable and more manipulable. In particular, the short time horizon for the agent makes it more important that the principal have a forward-looking performance measure in order to motivate the agent to be "long term" in his thinking. However, the short time horizon for the agent also means he has greater opportunities to avoid having to settle up if he misleads the principal about future prospects.

Multi-period models with sequential actions and the arrival of information between decisions are also necessary to address problems of (i) dynamically adjusting budgets and targets over time, (ii) motivating agents to collect and process information versus motivating them to use it appropriately, and (iii) motivating agents to acquire the proper amount of a productive resource versus motivating efficient allocation and use of the resource once it has been acquired. Multi-period models are also necessary to analyze learning over time. That is, what information can principals and agents collect to decide whether their strategies are "working" and how do they use information to adjust them over time? In would be especially interesting to examine agency applications of the learning model explored in Dye (1998), where the agent's operating decisions themselves affect the information that is generated at the end of the period. The principal and agent must trade off the benefits of "experimentation" to help make decisions that are better in the long run versus the cost of making decisions that are in the short run nonoptimal. This tradeoff is likely to be viewed differently between the two parties when the agent has a shorter time horizon than the principal.

There are significant challenges on both the "art" of modeling as well as the technical horsepower dimensions of doing theoretical research to address these issues. However, in order for agency theorists to continue to make substantive

contributions to accounting, these are the issues to be faced and the challenges to be overcome.

References

Alles, M., Datar, S., 1998. Strategic transfer pricing. Management Science 44 (4), 451–461.

Antle, R., 1982. The auditor as an economic agent. Journal of Accounting Research 488–500.

Antle, R., Demski, J., 1988. The controllability principle in responsibility accounting. The Accounting Review 700–718.

Antle, R., Eppen, G., 1985.Capital rationing and organizational slack in capital budgeting. Management Science 163–174.

Antle, R., Fellingham, J., 1995. Information rents and preferences among information systems in a model of resource allocation. Journal of Accounting Research 41–63.

Antle, R., Fellingham, J., 1997. Models of capital investment with private information and incentives: a selective review. Journal of Business Finance and Accounting 887–908.

Antle, R., Smith, A., 1986. An empirical investigation of the relative performance evaluation of corporate executives. Journal of Accounting Research 24 (1), 1–39.

Arnott, R., Stiglitz, J., 1988. Randomization and asymmetric information. Rand Journal of Economics 344–362.

Arrow, K., 1985. The economics of agency. In: Pratt, J., Zeckhauser, R. (Eds.), Principals and Agents: The Structure of Business. Harvard Business School Press, Boston, MA.

Arya, A., Glover, J., Sivaramakrishnan, K., 1997. The interaction between decision and control problems and the value of information. The Accounting Review 561–574.

Arya, A., Glover, J., Sunder, S., 1998. Earnings management and the revelation principle. Review of Accounting Studies 3 (1 and 2), 7–34.

Baiman, S., 1982. Agency research in managerial accounting: a survey. Journal of Accounting Literature 154–213.

Baiman, S., 1990. Agency research in accounting: a second look. Accounting Organizations and Society 341–371.

Baiman, S., Demski, J., 1980. Economically optimal performance evaluation and control. Journal of Accounting Research Supplement 184–220.

Baiman, S., Evans III, J.H., 1983. Pre-decision information and participative management control system. Journal of Accounting Research 371–395.

Baiman, S., Noel, J., 1982. Noncontrollable costs and responsibility accounting. Journal of Accounting Research 184–220.

Baiman, S., Rajan, M., 1994. On the design of unconditional monitoring systems in agencies. The Accounting Review 217–229.

Baiman, S., Rajan, M., 2000. Incentive issues in inter-firm relationships. Working paper, University of Pennsylvania.

Baiman, S., Rajan, M., The role of information and opportunism in the choice of buyer–supplier relationships. Journal of Accounting Research, Supplement 2001, forthcoming.

Baiman, S., Sivaramakrishnan, K., 1991. The value of private pre-decision information in a principal–agent context. The Accounting Review 747–766.

Baiman, S., Verrecchia, R., 1995. Earnings and price-based compensation contracts in the presence of discretionary trading and incomplete contracting. Journal of Accounting and Economics 93–121.

Baiman, S., Evans, H., Noel, J., 1987. Optimal contracts with a utility-maximizing auditor. Journal of Accounting Research 217–244.

Baiman, S., Evans, J., Nagarajan, N., 1991. Collusion in auditing. Journal of Accounting Research 1–18.

Baker, G., 1992. Contracts and performance measurement. Journal of Political Economy 100 (3), 598–614.

Baldenius, T., 2000. Intrafirm trade, bargaining power, and specific investments. Review of Accounting Studies 27–56.

Baldenius, T., Reichelstein, S., Sahay, S., 1999. Negotiated versus cost-based transfer pricing. Review of Accounting Studies 67–92.

Banker, R., Datar, S., 1989. Sensitivity, precision, and linear aggregation of signals for performance evaluation. Journal of Accounting Research 27 (1), 21–39.

Banker, R., Thevaranjan, A., 2000. Goal congruence and evaluation of performance measures. Working paper, University of Texas at Dallas.

Barclay, M., Gode, D., Kothari, S.P., 2000. The advantages of using earnings for compensation: matching delivered performance. Working paper, University of Rochester.

Bushman, R., Indjejikian, R., 1993a. Accounting income, stock price, and managerial compensation. Journal of Accounting and Economics 3–23.

Bushman, R., Indjejikian, R., 1993b. Stewardship value of 'distorted' accounting disclosures. The Accounting Review 765–782.

Bushman, R., Indjejikian, R., Penno, M., 2000. Private pre-decision information, performance measure congruity and the value of delegation. Working paper, University of Michigan.

Bushman, R., Indjejikian, R., Smith, A., 1995. Aggregate performance measures in business unit manager compensation: The role of intrafirm interdependencies. Journal of Accounting Research 33 (Suppl.), 101–128.

Christensen, J., 1981. Communication in agencies. Bell Journal of Economics 661–674.

Christensen, P., Feltham, G., Wu, M., 2000. Cost of capital in residual income performance evaluation. Working paper, University of British Columbia.

Datar, S., Kulp, S., Lambert, R., 2000. Balancing performance measures. Journal of Accounting Research, forthcoming.

Demski, J., 1981. Cost allocation games. In: Moriarity, S. (Ed.), Joint Cost Allocations. Center for Economic and Management Research University of Oklahoma, Norman, OK, pp. 142–173.

Demski, J., Dye, R., 1999. Risk, return, and moral hazard. Journal of Accounting Research 27–56.

Demski, J., Frimor, H., 1999. Renegotiation and income smoothing in agencies. Journal of Accounting Research 187–214.

Demski, J., Sappington, D., 1987. Delegated expertise. Journal of Accounting Research 68–90.

Demski, J., Sappington, D., 1993. Sourcing with unverifiable performance information. Journal of Accounting Research 1–20.

Demski, J., Patell, J., Wolfson, M., 1984. Decentralized choice of monitoring systems. The Accounting Review 16–34.

Dutta, S., Reichelstein, S., 1999. Asset valuation and performance measurement in a dynamic agency setting. Review of Accounting Studies.

Dutta, S., Reichelstein, S., 2000. Controlling investment decisions: hurdle rates and intertemporal cost allocation. Working paper, University of California at Berkelely.

Dye, R., 1984. The trouble with tournaments. Economic Inquiry 147–149.

Dye, R., 1985. Costly contract contingencies. International Economic Review 233–250.

Dye, R., 1986. An optimal monitoring policy in agencies. The Rand Journal of Economics 339–350.

Dye, R., 1988. Earnings management in an overlapping generations model. Journal of Accounting Research 195–235.

Dye, R., 1992. Relative performance evaluation and project selection. Journal of Accounting Research 27–52.

Evans, J., Sridhar, S., 1996. Multiple control systems, accrual accounting and earnings management. Journal of Accounting Research 45–66.

Fellingham, J., Kwon, Y., Newman, D.P., 1984. Ex ante randomization in agency models. Rand Journal of Economics 290–301.

Fellingham, J., Newman, D., Suh, Y., 1985. Contracts without memory in multiperiod agency models. Journal of Economic Theory 37 (2), 340–355.

Feltham, G., Wu, M., 2000. Public reports, information acquisition by investors, and management incentives. Review of Accounting Studies, forthcoming.

Feltham, G., Wu, M., 2001. Incentive efficiency of stock versus options. Review of Accounting Studies 6 (1), 7–28.

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

Fershtman, C., Judd, K., 1987. Equilibrium incentives in oligopoly. American Economic Review.

Fischer, P., 1999. Managing employee compensation risk. Review of Accounting Studies 45–60.

Fischer, P., Verrecchia, R., 2000. Reporting bias. The Accounting Review 229–245.

Fudenberg, D., Tirole, J., 1990. Moral hazard and renegotiation in agency contracts. Econometrica 1279–1320.

Gigler, F., Hemmer, T., 1998. On the frequency, quality, and informational role of mandatory financial reports. Journal of Accounting Research Supplement 117–160.

Gjesdal, F., 1981. Accounting for stewardship. Journal of Accounting Research 208–231.

Grossman, S., Hart, O., 1983. An analysis of the principal–agent model. Econometrica 7–45.

Hansen, S., Magee, R., 1993. Capacity cost and capacity allocation. Contemporary Accounting Research 635–660.

Harris, M., Kriebel, C., Raviv, A., 1982. Asymmetric information, incentives, and intrafirm resource allocation. Management Science 604–620.

Hart, O., Holmstrom, B., 1986. The theory of contracts. In: Bewlely, T. (Ed.), Advances in Economic Theory. Cambridge University Press, Cambridge.

Healy, P., 1985. The effect of bonus schemes on accounting decisions. Journal of Accounting and Economics 85–107.

Hemmer, T., 1996. On the design and choice of "modern" management accounting measures. Journal of Management Accounting Research 8, 87–116.

Hemmer, T., Kim, O., Verrecchia, R., 1999. Introducing convexity into optimal compensation contracts. Journal of Accounting and Economics 28 (3), 307–327.

Holmstrom, B., 1979. Moral hazard and observability. Bell Journal of Economics 74–91.

Holmstrom, B., Milgrom, P., 1987. Aggregation and linearity in the provision of intertemporal incentives. Econometrica 303–328.

Holmstrom, B., Milgrom, P., 1991. Multi-task principal–agent analyses: incentive contracts, asset ownership, and job design. Journal of Law, Economics, and Organization 24–52.

Indjejikian, R., 1999. Performance evaluation and compensation research: an agency theory perspective. Accounting Horizons 147–157.

Indjejikian, R., Nanda, D., 1999. Dynamic incentives and responsibility accounting. Journal of Accounting and Economics 27 (2), 177–201.

Ittner, C., Larcker, D., Rajan, M., 1997. The choice of performance measures in annual bonus contracts. The Accounting Review 72 (2), 231–255.

Janakiraman, S., Lambert, R., Larcker, D., 1992. An empirical investigation of the relative performance. Journal of Accounting Research 53–69.

Jewitt, I., 1988. Justifying the first-order approach to principal–agent problems. Econometrica 1177–1190.

Kaplan, R., Norton, D., 1992. The balanced scorecard: Measures that drive performance. The Harvard Business Review, January–February, 71–79.

Kaplan, R., Norton, D., 1993. Putting the balanced scorecard to work. The Harvard Business Review, September–October, 134–147.

Kim, S., Suh, Y., 1991. Choice of accounting information system. Journal of Accounting Research 386–396.

Kim, O., Suh, Y., 1993. Incentive efficiency of compensation based on accounting and market performance. Journal of Accounting and Economics 25–53.

Kirby, A., Reichelstein, S., Sen, P., Paik, T., 1991. Participation, slack, and budget-based performance evaluation. Journal of Accounting Research 109–128.

Kreps, D., 1990. Moral hazard and incentives. In: A Course in Microeconomic Theory. Princeton University Press, Princeton, NJ, pp. 577–617 (Chapter 16).

Kulp, S., 2000. Asymmetric information in vendor managed inventory systems. Working paper, Harvard Business School.

Lambert, R., 1983. Long-term contracts and moral hazard. The Bell Journal of Economics 14 (2), 441–452.

Lambert, R., 1985. Variance analysis in agency settings. Journal of Accounting Research 633–647.

Lambert, R., 1986. Executive effort and the selection of risky projects. The Rand Journal of Economics 77–88.

Lambert, R., 1993. The use of accounting and security price measures of performance in managerial compensation contracts: a discussion. Journal of Accounting and Economics 16 (1–3), 101–123.

Lambert, R., Larcker, D., 1987. An analysis of the use of accounting and market measures of performance in executive compensation contracts. Journal of Accounting Research 25 (Suppl.), 85–125.

Magee, R., 1988. Variable cost allocations in a principal–agent setting. The Accounting Review 42–54.

Melumad, N., Reichelstein, S., 1989. The value of communication in agencies. Journal of Economic Theory.

Merchant, K., 1987. Why firms disregard the controllability principle. In: Bruns, W., Kaplan, R. (Eds.), Accounting and Management: Field Study Perspectives. Harvard Business School Press, Boston, MA, pp. 316–338.

Meth, B., 1996. Reduction of outcome variance: optimality and incentives. Contemporary Accounting Research 309–328.

Milgrom, P., 1981. Good news and bad news: representation theorems and applications. Bell Journal of Economics 380–391.

Milgrom, P., Roberts, J., 1992. Economics, Organization, and Management. Prentice-Hall, Englewood Cliffs, NJ.

Miller, B., Buckman, G., 1987. Cost allocation and opportunity costs. Management Science 626–639.

Mirrlees, J., 1974. Notes on welfare economics, information, and uncertainty. In: Balch, M., McFadden, D., Shih-Yen Wu (Eds.), Essays on Economic Behavior under Uncertainty. North-Holland Publishing Co., Amsterdam.

Myerson, R., 1979. Incentive compatibility and the bargaining problem. Econometrica 61–74.

Narayanan, V.G., Davila, A., 1998. Using delegation and control systems to mitigate the trade-off between the performance-evaluation and belief-revision uses of accounting signals. Journal of Accounting, Economics 25 (3), 255–282.

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

Penno, M., 1984. Asymmetry of pre-decision information and managerial accounting. Journal of Accounting Research 177–191.

Pratt, J., Zeckhauser, R., 1985. Principals and agents: an overview. In: Pratt, J., Zeckhauser, R. (Eds.), Principals and Agents: The Structure of Business. Harvard Business School Press, Boston, MA ((Chapter 1)).

Prendergast, C., 1999. The provision of incentives in firms. Journal of Economic Literature 7–63.

Rajan, M., 1992. Cost allocation in multi-agent settings. The Accounting Review 527–545.

Rajan, M., Sarath, B., 1997. The value of correlated signals in agencies. The Rand Journal of Economics 150–167.

Reichelstein, S., 1993. Constructing incentive schemes for government contracts an application of agency theory. The Accounting Review October 712–731.

Reichelstein, S., 1997. Investment decisions and managerial performance. Review of Accounting Studies 2 (2), 157–180.

Reichelstein, S., 2000. Providing managerial incentives: cash flows versus accrual accounting. Journal of Accounting Research 243–269.

Rogerson, W., 1985a. The first order approach to principal–agent problems. Econometrica 1357–1368.

Rogerson, W., 1985b. Repeated moral hazard. Econometrica 53 (1), 69–76.

Rogerson, W., 1997. Inter-temporal cost allocation and managerial incentives. Journal of Political Economy 770–795.

Ronen, J., McKinney, G., 1970. Transfer pricing for divisional autonomy. Journal of Accounting Research 99–112.

Sahay, S., 1997. Studies in the theory of transfer pricing. Ph.D. Dissertation, University of California at Berkeley.

Sansing, R., 1999. Relationship-specific investments and the transfer pricing paradox. Review of Accounting Studies 119–134.

Sloan, R., 1993. Accounting earnings and top executive compensation. Journal of Accounting and Economics 16 (1–3), 55–100.

Smith, M., 1999. Tax and incentive trade-offs in transfer pricing. Working paper, Duke University.

Stocken, P., Verrecchia, R., 1999. Performance monitoring and financial disclosure choice. Journal of Institutional and Theoretical Economics 155, 214–232.

Thevaranjan, A., Joseph, K., Srinivasan, D., 2000. Managerial myopia and non-financial measures: The case of customer satisfaction mitigating hard selling. Working paper, Syracuse University.

Verrecchia, R., 1986. Managerial discretion in the choice among financial reporting alternatives. Journal of Accounting and Economics 175–196.

Verrecchia, R., 2000. Essays on disclosure. Journal of Accounting and Economics, forthcoming.

Weitzman, M., 1976. The new Soviet incentive model. Bell Journal of Economics 251–257.

Wilson, R., 1968. On the theory of syndicates. Econometrica 119–132.

Young, R., 1986. A note on economically optimal performance evaluation and control systems: the optimality of two-tailed investigations. Journal of Accounting Research 231–240.