RICE UNIVERSITY

Toward an Accounting-Centric Principal-Agent Framework: Theory and Some Applications

by

Amoray Riggs-Cragun

A THESIS SUBMITTED
IN PARTIAL FULFILLMENT OF THE
REQUIREMENTS FOR THE DEGREE

Doctor of Philosophy

APPROVED, THESIS COMMITTEE:

Thomas Hemmer, Chair
Houston Endowment Professor of Accounting

David De Angelis
Assistant Professor of Finance

Ted Loch-Temzelides
Professor of Economics

Houston, Texas

March, 2019

Toward an Accounting-Centric Principal-Agent Framework: Theory and Some Applications

Amoray Riggs-Cragun

Thesis: Doctor of Philosophy
Accounting, Jones Graduate School of Business
Rice University, Houston, Texas (March 2019)

ABSTRACT

I develop a principal-agent framework in which agents take actions that influence a firm's general ledger accounts. My framework is tractable and facilitates optimal closed-form solutions without assuming preferred actions or imposing exogenous restrictions on contractual form. I apply the framework to several settings and a consistent theme emerges: it is cheaper to motivate efficiency improvements than revenue growth. This is driven by a seemingly trivial property of bookkeeping, that general ledger accounts are bounded below by zero. I find that this accounting feature influences task allocation, organizational design and optimal aggregation rules. My analyses produce many readily-testable predictions and can help explain empirical regularities; for example, I predict that the low pay-performance sensitivities empirically observed in loss firms may be driven in part by life cycle. My findings demonstrate that when studying the optimal use of accounting information, properties of the double-entry systems which generate that information should be taken seriously.

ACKNOWLEDGMENTS

I am deeply grateful to my adviser, Thomas Hemmer, for his mentorship, encouragement and guidance. He has been the sole champion of my doctoral studies at Rice and has advocated valiantly on behalf of my education. His insights were invaluable as I developed this thesis, and his mentorship has shaped me as a scholar. I feel incredibly fortunate to have been his student, and my research, career trajectory and life are better for it.

I would like to thank the other members of my committee, David De Angelis and Ted Loch-Temzelides, for useful advice and feedback. I offer my thanks to the many other Rice professors in accounting, finance and economics who have contributed to my education. I am also grateful to my current and former fellow doctoral students for many enjoyable years of comradery, learning, discussion and debate.

I express a heartfelt thanks to my husband, Nathan, who has supported my education unquestioningly and has helped me in ways that go far beyond the normal expectations of a spouse. I am grateful to have supportive parents who raised me to love learning. I thank my mother, Marti, who is my constant cheerleader and who generously spent many weeks helping us with our daughter as I worked on this thesis. I thank my father, Russ, and his partner, Jesse, for their support and encouragement throughout my PhD program, including their gracious hospitality when I spent summer months doing research in their lovely Oregon home.

CONTENTS

1

INTRODUCTION

General ledger accounts are the basic building blocks of accounting performance measures. Accountants have specialized knowledge about these accounts and the double-entry systems in which they reside. This expertise appears underutilized in theoretical accounting research, which tends to model the entire accounting report with a single performance measure - earnings - without regard to its underlying components or features of the accounts which comprise them. Because earnings is an aggregate measure, managers can only influence it by affecting one or more of its components. If those components respond differently to managerial effort, then optimal incentives should depend on the component(s) being influenced.

In this thesis, I develop a principal-agent framework in which managers stochastically influence a firm's accounts, which can serve as performance measures themselves or can be aggregated to form other measures, such as earnings. I make my framework tractable by imposing Poisson distributions, square-root utility preferences and context-specific cost of effort and profit functions. I demonstrate the features of my framework in five applications focused on two simple earnings-increasing activities: cost cutting and revenue growth. I find that the zero bound on revenue and expense accounts affects optimal contracting and production under moral hazard, and consequently has broad organizational ramifications, influencing optimal aggregation rules, task allocation, and organizational design. In each application, I discuss the empirical implications of my findings.

In my framework, an accounting system records a firm's transactions in Poisson-distributed accounts. Consistent with accounts in double-entry systems, the Poisson distribution is unbounded above and bounded below by zero. Poisson-distributed variables are also amenable to aggregation and thus well-suited for modeling the construction of accounting reports. The firm's risk-neutral owner hires agent(s) to take action(s) that increase expected firm value and stochastically influence one or more of the firm's general ledger accounts. I model this influence by making each account's Poisson parameter a function of managerial actions. A given action may increase some accounts, decrease others or affect how accounts are related, allowing for rich representations of the ways managers influence firms. My framework technology is quite flexible; the action set, account set, and the functions by which actions affect each account can all be tailored to the specific setting being studied.

Admittedly, relying on the Poisson specification is not without loss of generality. But what I gain from this parametric structure, in combination with the other assumptions of my model, is tractability. Intractability has long been an impediment to studying accounting issues in an agency theory context, as the general nature of the principal-agent (P-A) model leaves few degrees of freedom for investigating incentive issues that go beyond the optimal design of the contract.

One popular approach for addressing the tractability problem is what's known as the LEN model, which assumes that an agent with negative exponential utility controls the mean (and only the mean) of normally-distributed performance measures and that her contract is linear in those performance measures. These assumptions clear away the cumbrous machinery of the general P-A problem, making the LEN model algebraically accessible even in complex settings. However, when LEN's linear contract restriction does not align with truly optimal contracts, it can produce results that are based on the *sub*optimal use of information. To its credit, the LEN model has been extremely influential, and its widespread use demonstrates that imposing structure on the P-A problem can facilitate the tractability needed to study applied issues involving information. While I take a similar approach of adding structure to the general problem, I do so without restricting the contract to take a particular form, thereby avoiding the problematic possibility of drifting into a third-best analysis.

The optimal, unrestricted contracts that emerge from my framework are markedly simple. A square root utility function, in combination with the linear likelihood ratios associated with Poisson-distributed accounts, makes the optimal contract linear in account balances in utility space (and convex in cash space). I show that the optimal contractual weight on a given account depends on the agent's cost function and on the setting-specific technology by which managerial actions affect the account. I find that optimal aggregation rules are more aggressive (i.e. they weight revenues more heavily than expenses) when a firm has been reporting extreme losses, has expense accounts loaded with fixed costs, or competes more on cost than on product differentiation. I also find that optimal weighting rules are more conservative when cost of effort functions are more convex.

One benefit of my framework is the ease by which it allows one to study the effects of moral hazard on desired actions. Many applied agency papers deal with optimal contracts for a *given* action; for example, papers following Gigler and Hemmer (2001) assume a binary action space in which the agent can choose either high or low effort. Because these models assume a certain preferred action *ex ante*, they are rather limited in their ability to speak to issues involving production, optimal task assignment or job design.

I study two primary types of actions: revenue-increasing actions and cost-cutting actions. Very consistently, I find that cost-cutting actions are cheaper to implement than revenue-increasing actions. This occurs because, due to the lower bound of zero on the accounts, managerial action affects the variance of the accounts in opposite ways: revenue growth increases the variance of revenues, while cost reductions decrease the variance of expenses. To see this, visualize a revenue distribution that is unimodal, bounded below by zero and unbounded above. As the manager takes actions that increase expected revenues, the distribution's mass is spread rightward over a wider area and the account's variance increases. Therefore, a higher equilibrium action injects more variance into the performance measure and thus more risk into the agent's contract. Compensating the agent for this added risk makes the higher revenue-increasing action more expensive to implement. In contrast, cost-cutting actions decrease the expected value of the expense account, compressing its mass tighter against the lower bound of zero and decreasing its variance, which in turn reduces the risk in the agent's contract.

This phenomenon - that expense-cutting actions are inherently less risky than revenue-increasing actions - arises because account mean and variance move together. This comes along with my Poisson specification, because the mean and variance of the Poisson are equal, but it is not particular to the Poisson. Mean and variance appear to move together for any distribution which (1) is bounded below and unbounded above, and (2) responds to mean-shifting managerial action in accordance with the monotone likelihood ratio condition.

The assumption that account mean and variance move together seems reasonably consistent with many real-world settings. A manager hired to increase revenues will likely have to take bold or risky actions to do so, as any risk-free opportunities for revenue growth will have already been exhausted by competition. For expenses, the argument is most obvious in the limit: eliminating all production costs would eliminate the variance of those costs as well.

While I argue that account mean and variance move together in many real-world operational settings, the assumption does not hold universally. For example, reducing the expected value of income tax expense may require taking riskier tax positions. My framework cannot speak to these settings; it applies only to settings in which mean-increasing actions are also variance-increasing.

I conceptualize revenues as $price * volume$ and examine settings in which the manager is tasked with both curbing costs and increasing either sales volume or selling price. I find that the degree to which cost cutting is emphasized over growth is stronger in volume-focused firms relative to price-focused firms. This difference arises because sales volume affects expenses but selling price does not. Consequently, providing incentives for volume growth creates indirect incentives for cost cutting because it is in the agent's interest to combat the associated increase in expense account variance. I also find that optimal aggregation rules are aggressive for volume-focused firms and neutral for price-focused firms. I discuss the empirical implications of these findings and suggest a textual analysis approach for identifying volume- versus price-focused firms.

One application of my framework presents a particularly stark version of the finding that cost cutting is cheaper to implement than revenue growth. I show that when the restriction of a square root utility function is lifted and replaced with a general concave utility function that is unbounded below, implementing the cost-cutting action is *free* from a risk-sharing perspective. This is achieved through a penalty scheme a la Mirrlees (1999), in which the manager is paid the first-best wage if expenses fall below some threshold and incurs a penalty if expenses exceed that threshold. The scheme works because bad expense outcomes are unbounded above (assuming unlimited credit or borrowing), so as expense realizations move toward the extreme right tail of the distribution, the principal becomes increasingly confident that the manager shirked. For a given threshold, the principal threatens a penalty so painful that the manager *never* shirks, thereby guaranteeing the first-best action. The threshold and penalty can be set increasingly high until, in the limit, the manager takes the first-best action, earns the first-best wage and never incurs the penalty.

Penalty schemes are not efficient for implementing revenue-increasing actions. This is because bad revenue outcomes are *bounded*. There is a worst possible revenue outcome - specifically, zero dollars - making it impossible for the principal to detect shirking with adequate certainty to approximate the first-best solution.

The finding that penalty schemes are efficient for cost cutting but not revenue growth has empirical implications for pay-performance sensitivities (PPS), which I discuss at length in section 3.5. If one is willing to believe that at least *some* firms employ penalty schemes to implement cost cutting, then on average, agents tasked with cutting costs should have less variable compensation than those tasked with revenue growth. This allows for PPS predictions across cross sections in which managers are presumed to be tasked more heavily with cost cutting versus growth. For example, it seems reasonable to assume that the emphasis on cost cutting relative to growth is higher for Chief Operations Officers (COOs) than for Chief Marketing Officers (CMOs). My findings would then suggest that the sensitivity of COO compensation to operating expenses should be lower than the sensitivity of CMO compensation to revenues. Notice that this prediction deals with sensitivities to account-level performance (i.e. expenses for the COO and revenues for the CMO) rather than aggregate earnings performance; I distinguish between *account-PPS* and *earnings-PPS* and make predictions about each.

I predict that earnings-PPS will be low (or even negative) in startup firms, increase as firms grow and reach maturity, and decrease as firms enter decline. This suggests a new explanation for the well-documented phenomenon of lower PPS in loss firm-years relative to profitable firm-years: to the extent that losses are more common during the startup and decline phases, the low PPS in loss years may be explained in part by life cycle.

Finally, I study how incentives for revenue growth and cost cutting affect organizational design. In a setting with N products and N agents, I study job diversification and optimal team size for a revenue center and a cost center. Members of a sales or cost-cutting team are compensated on the total sales or production costs, respectively, of the products assigned to their team. In choosing whether to increase team size, the principal faces a trade-off between the beneficial synergies of job diversification and the cost of compensating each agent for the added risk in her contract produced by assessing her performance on measures affected by an increasingly large team. Because cost-cutting actions are variance reducing, the added contractual risk from increasing team size is smaller for cost-cutting tasks than for sales tasks. This results in optimal team size being larger in cost centers than revenue centers.

I make three primary contributions. First, I design an analytical representation of accounting that models economic activity at the account level and mirrors real-world attributes of general ledger accounts (section 2.1). While I employ this structure in an agency context, disclosure theorists may find it useful as well. Second, I provide a set of assumptions that make the principal-agent model algebraically tractable without assuming preferred actions or imposing exogenous restrictions on contractual form (section 2.2). This model specification should be particularly useful for studying settings in which performance measures are accounting-generated or have distributional features similar to the Poisson (such as being bounded below by zero and unbounded above). Third, I find that agency costs tend to be lower for cost cutting than for revenue growth (section 3.1), which to my knowledge has not been demonstrated in the agency literature. My findings show that the effects of this phenomenon ripple throughout the organization, interacting with firm strategy (section 3.2), team formation (section 3.3), optimal aggregation rules (sections 3.2 and 3.4), and pay-performance sensitivities (section 3.5). I believe the applications presented only scratch the surface of what this framework can address; section 4 discusses my plans for future work. Overall, I aim to demonstrate that an account-oriented approach to modeling accounting information can provide new insights into longstanding economic questions.

2

FRAMEWORK

Consider the following single-period model. A firm's accounting system tracks its economic activity by recording meaningful events and transactions in accounts which act as random variables.There are n such accounts, with ending balances represented by $\vec{x} \equiv (x_1, ..., x_n)$. The firm's risk-neutral owner hires a risk-averse manager to exert continuous effort along m unobservable tasks, denoted $\vec{a} \equiv (a_1, ..., a_m)$, with $a_i \in \mathbb{R}^+$ for all i. The agent's actions impose a personal cost on the agent of $c(\vec{a})$, where $c(\cdot)$ is increasing and convex. The agent's reservation utility is $\bar{U}$ and her utility from compensation, $u(\cdot)$, is increasing, concave and is additively separable from her cost of effort. The agent's actions parametrically influence the firm's accounts ($\vec{x}$), which are used as performance measures to evaluate the agent. Let $\Pi(\vec{a})$ be the principal's gross expected payoff from hiring the manager to take action vector $\vec{a}$. The principal's objective is to maximize her expected net profit by choosing a set of actions ($\vec{a}$) as well as a compensation scheme $s(\vec{x})$ to implement those desired actions. The principal's problem is shown below.

$$
\begin{aligned}
&\max_{s(\vec{x}),\vec{a}} && \Pi(\vec{a}) - E[s(\vec{x})|\vec{a}] && \text{(OBJ)} \\
&\text{subject to} && E[u(s(\vec{x}))|\vec{a}] - c(\vec{a}) \geq \bar{U} && \text{(IR)} \\
&\text{and} && \vec{a} \in \operatorname*{argmax}_{\tilde{a}} \{E[u(s(\vec{x}))|\tilde{a}] - c(\tilde{a})\} && \text{(IC)}
\end{aligned}
\tag{2.1}
$$

The individual rationality (IR) constraint ensures that the agent is willing to accept the contract, and the incentive compatibility (IC) constraint ensures that the agent takes the appropriate actions by making $\vec{a}$ in the agent's best interest given the contract. Notice that even in the single action case in which $\vec{a} = a \in \mathbb{R}$, (IC) is a *set* of constraints and represents infinitely many pairwise comparisons of the agent's expected utility under all possible actions. This is a major obstacle to tractable analysis of the problem, as it is difficult to determine which constraints bind. The *first-order approach* (FOA) is a commonly used method for simplifying the IC constraint set; it replaces (IC) with the first-order necessary condition(s) from the agent's unconstrained maximization problem.

Even with the first-order approach, the general nature of the principal-agent model makes it difficult to address specific or applied issues because characterizing the optimal contract tends to burn up most of the researcher's degrees of freedom. The popular LEN framework addresses this technical problem by imposing a particular structure on the P-A problem. LEN requires that $s(\cdot)$ is linear in performance measures, that the agent has (multiplicatively separable) negative exponential utility and that signals are normally distributed; additionally, most LEN models assume that the agent's cost of effort function is quadratic. This combination of assumptions provides tractability even in complex settings, which has facilitated the study of many interesting organizational and economic issues in the literature. However, LEN's tractability comes at a cost: the ex ante restriction of linear contracts might not align with truly optimal contracts.

In this section, I develop a principal-agent framework oriented around two primary objectives. First, I aim to replicate LEN's convenience and algebraic accessibility without its potential downside. I do this by, like LEN, taking a parametric approach that imposes a specific structure on the P-A program, but I do so without placing ex ante restrictions on the contractual form. Second, the parametric assumptions I impose are intended to capture fundamental attributes of accounting information systems, making it well-suited for addressing settings with accounting-generated performance measures like the one described above. Section 2.1 details these distributional assumptions; this is the most important piece of the framework because it is what makes it *accounting-centric*. Section 2.2 shows how the problem can be made more tractable by imposing additional structure on (in order of importance) the utility function $u(s)$, the cost of effort function $c(\vec{a})$ and the profit function $\Pi(\vec{a})$.

2.1 A Parametric Representation of General Ledger Accounts

While certainly convenient, the FOA is unfortunately not always valid (Mirrlees (1999)). Jewitt (1988) put forth a set of sufficient conditions for validity of the FOA, and I adhere to his conditions in building my framework.[1] This limits the set of possible distributions I can use to represent the account technology. Jewitt identifies three well-known distributions that satisfy his conditions: the Chi-squared, gamma, and Poisson. Of these three candidates, the Poisson best suited to model accounting because it is most amenable to addition and subtraction, which is at the heart of basic accounting mechanics. Additionally, the Poisson distribution mirrors the basic properties of double-entry accounts: it is bounded below by zero and unbounded above.

For these reasons, I assume that accounts in the firm's double-entry system are Poisson-

1. A different set of sufficient conditions were provided by Mirrlees (1979) and later proven correct by Rogerson (1985). These conditions included the monotone likelihood ratio property (MLRP) and the convexity of the distribution function condition (CDFC). The CDFC is extremely difficult to satisfy and rules out almost all known probability distributions, so I use Jewitt's less restrictive conditions rather than the Mirrlees-Rogerson conditions.

distributed. Let x_j represent account j and be distributed as follows.

$$x_j \sim Poisson(\bar{x}_j), \text{ with } \bar{x}_j \equiv \hat{x}_j + f_j(\vec{a}), \tag{A1}$$

The Poisson distribution has one parameter; I call that parameter $\bar{x}_j$ and separate it into two additive components. The first, $\hat{x}_j$, is exogenous and represents the expected value of x_j if the agent takes no action. The second, $f_j(\vec{a})$, is the effect of the agent's actions on the expected value of x_j.

The distribution of any Poisson-distributed account x is unimodal, positively skewed and has the following probability mass function.

$$Pr(x|\bar{x}) = \frac{e^{-\bar{x}}\left(\bar{x}^x\right)}{x!}, \quad x = 0, 1, 2, ... \tag{2.2}$$

The first four moments of account x are as follows.

$$E(x) = \bar{x} \tag{2.3}$$

$$Var(x) = \bar{x} \tag{2.4}$$

$$Skewness(x) = \frac{1}{\sqrt{\bar{x}}} \tag{2.5}$$

$$Kurtosis(x) = \frac{1}{\bar{x}} \tag{2.6}$$

Notice that all moments of the Poisson distribution are affected by the manager's effort; if the manager influences an account's expected value, she influences all other moments as well.

The sum of Poisson-distributed random variables is also Poisson-distributed; if x_s is the sum of $x_1 \sim Poisson(\bar{x}_1)$ and $x_2 \sim Poisson(\bar{x}_2)$, then $x_s \sim Poisson\left(\bar{x}_1 + \bar{x}_2\right)$. The difference between two Poisson-distributed random variables follows the two-parameter *Poisson-difference distribution* or *Skellam distribution* (Skellam 1946). Distributions of the sums or differences of Skellam-distributed accounts are also Skellam-distributed.[2]

With well-defined sums and differences of account balances, one could use this framework to model the financial statements by appropriately classifying and aggregating general ledger accounts. No matter how the aggregation is done, every line item will follow either a Poisson or Skellam distribution. This technology is illustrated in Figure 1, which shows a simple income statement in which (Poisson-distributed) revenue accounts aggregate to Poisson-distributed total revenues, (Poisson-distributed) expense accounts aggregate to Poisson-distributed total expenses, and total revenues and total expenses are differenced to arrive at Skellam-distributed earnings.

Let $LR_j^i \equiv \frac{\frac{\partial}{\partial a_i}(Pr(x_j|\vec{a}))}{Pr(x_j|\vec{a})}$ denote the likelihood ratio for account x_j in reference to action $a_i \in \vec{a}$. The following observation provides this likelihood ratio.[3]

Observation 1. *The likelihood ratio for any Poisson-distributed account x_j and action a_i is given as*

$$LR_j^i \equiv \frac{\frac{\partial}{\partial a_i}\left(Pr(x_j|\vec{a})\right)}{Pr(x_j|\vec{a})} = \left(\frac{x_j - \bar{x}_j}{\bar{x}_j}\right) f_j^i, \quad \textit{where } f_j^i \equiv \frac{\partial f_j(\vec{a})}{\partial a_i}. \tag{2.7}$$

The likelihood ratio has an expected value of zero and the following variance.

$$Var(LR_j^i) = \frac{\left(f_j^i\right)^2}{\bar{x}_j} \tag{2.8}$$

The likelihood ratio is a product of the marginal influence of a_i on $E[x_j]$ times the scaled deviation of account x_j from its expectation given actions $\vec{a}$. Notice that the likelihood ratio is linear in x_j; this will play an important role as I develop tractability in the next section.

2.2 Tractability

One of Jewitt's (sufficient) conditions for validity of the FOA is that $u(s)$ is a concave transformation of $1/u'(s)$. The square root satisfies this condition, and it seems to be

2. Because the settings I investigate in this paper rely on Poisson- rather than Skellam-distributed accounts, I relegate details on the Skellam distribution to Appendix A.

3. All proofs are provided in Appendix B.

"less risk averse" than other utility functions that satisfy the condition. For example, the logarithimic utility function $ln(s)$ can be rearranged as $2ln\left(\sqrt{s}\right)$; therefore logarithimic utility is a concave transformation of square root utility and thus represents more risk aversion. Additionally, all power utility functions s^{α} with $\alpha \leq 1/2$ satisfy Jewitt's condition and have a higher constant relative risk aversion than the square root. Within the class of utility functions satisfying Jewitt's condition, the square root appears to be among the least risk averse.[4] With a risk-neutral principal, all second-best results are driven by the agent's risk aversion, so many utility functions within the class satisfying Jewitt's condition would produce more extreme versions of results obtained using the square root.

A square root utility function, in combination with a risk-neutral principal, makes the compensation contract quadratic in likelihood ratios (see for example Kim and Suh (1991)). Hemmer, Kim, and Verrecchia (2000) show that this gives the opportunity for closed-form solutions, particularly when likelihood ratios are linear in performance signals. Specifically, they show that a square-root utility function, in conjunction with the linear likelihood ratios of the gamma distribution, facilitates closed-form solutions. The Poisson distribution also exhibits linear likelihood ratios, as shown in Observation 1. Therefore, in pursuit of making my accounting-oriented framework tractable, I assign the agent a square root utility function:

$$u(s) = \sqrt{s}. \tag{A2}$$

Assumption (A2) allows me to rewrite program (2.1) as follows, where $c^i \equiv \frac{\partial}{\partial a_i} c(\vec{a})$ is the marginal cost of action a_i.

$$\begin{aligned} &\max_{s(\vec{x}),\vec{a}} && \Pi(\vec{a}) - E[s(\vec{x})|\vec{a}] && \text{(OBJ)} \\ &\text{subject to} && E[\sqrt{s}|\vec{a}] - c(\vec{a}) \geq \bar{U} && \text{(IR)} \\ &\text{and} && \tfrac{\partial}{\partial a_i} E[\sqrt{s}|\vec{a}] = c^i && (\text{IC}_i), \quad \text{for all } i \in (1, ..., m) \end{aligned} \tag{2.9}$$

4. The negative exponential utility function also satisfies Jewitt's condition, but its degree of risk aversion depends on the specified coefficient of absolute risk aversion, making direct comparison with the square root difficult.

The following lemma gives the optimal contract for this (m-action, n-account) program, where λ is the multiplier on the (IR) constraint and $\mu_1, \mu_2, ..., \mu_m$ are the respective multipliers on constraints (IC$_1$), (IC$_2$), ... , (IC$_m$).

Lemma 1. *The contract optimizing program* (2.9) *is characterized by*

$$2\sqrt{s} = \lambda + \sum_{i=1}^{m} \mu_i \sum_{j=1}^{n} f_j^i \left(\frac{x_j - \bar{x}_j}{\bar{x}_j} \right), \tag{2.10}$$

with multipliers

$$\lambda = 2\bar{U} + 2c(\vec{a}) \tag{2.11}$$

$$\textit{and } \mu_k = \frac{\det(\mathrm{D}_k)}{\det(\mathrm{D})} \textit{ for all } k \in (1, ..., m), \tag{2.12}$$

where D *is the* $m \times m$ *matrix with element* $d_{ik} \equiv \sum_{j=1}^{n} \frac{f_j^i f_j^k}{\bar{x}_j}$ *in row* i *and column* k*, and* D_k *is the matrix formed by replacing column* k *of* D *with* $\boldsymbol{c}' \equiv (c^1 c^2 ... c^m)^{\mathrm{T}}$.

To ensure that the square root utility function is well defined, I assume that contractual payments are bounded below by zero; that is, $s(\cdot) \geq 0$.With a lower bound on the left-hand side of contract (2.10), in some cases it is technically necessary to right-truncate certain account distributions so that the likelihood ratios in the contract are appropriately bounded below on the right-hand side of (2.10). The potential for likelihood ratios being unbounded below arises when an equilibrium action a_i decreases the expected value of account x_j; that is, when $f_j^i < 0$. Notice from equation (2.7) that $f_j^i < 0$ makes the likelihood ratio of account x_j with respect to action a_i decreasing in x_j, and because x_j is unbounded above and the likelihood ratio is linearly decreasing, LR_j^i is consequently unbounded below. In these cases, I assume that account x_j follows a *right-truncated* Poisson distribution with no probability mass above $x_j = R$: $x_j \sim \textit{RT-Poisson}(\bar{x}_j, R), x_j = 0, 1, ..., R$. The likelihood ratio of this account with respect to action a_i is

$$f_j^i \left(\frac{x_j}{\bar{x}_j} - \frac{\sum_{k=0}^{R} \frac{\bar{x}_j^{k-1}}{(k-1)!}}{\sum_{k=0}^{R} \frac{\bar{x}_j^{k}}{k!}} \right).$$

The second fraction inside the parentheses above rapidly approaches 1 as R increases (by $R = 25$, for example, the value of the quotient is .9999971). The likelihood ratio for a regular Poisson-distributed account is $f_j^i \left(\frac{x_j}{\bar{x}_j} - 1\right)$, and therefore the truncated likelihood ratio rapidly approaches the regular likelihood ratio as the truncation point is increased. With sufficiently high R the likelihood ratios are indistinguishable for $x_j < R$. Consequently, because contract 2.10 approximates a contract written over truncated Poisson accounts, and because the regular Poisson distribution is easier to work with than the truncated one, I use regular Poisson distributions in my analyses regardless of the sign of f_j^i.[5]

Lemma 1 shows that the optimal contract (2.10) in utility space is a linear combination of the firm's accounts, where the weight on account x_j is

$$\sum_{i=1}^{m} \mu_i f_j^i.$$

The multipliers $(\mu_1, \mu_2, ..., \mu_m)$ depend on the account technology and the agent's marginal cost of effort in each task, $(c^1, c^2, ..., c^m)$. In order to allow for closed-form solutions, it is therefore necessary to specify the agent's cost of effort function. In most of the applications I explore, the cost of effort function is not of primary interest and my cost of effort function specification is driven by tractability considerations. Specifically, I assume that $c(\vec{a})$ is linear and additively separable in all tasks.

$$c(\vec{a}) = \delta_1 a_1 + \delta_2 a_2 + ... + \delta_m a_m, \tag{A3}$$

where δ_i is the marginal cost of action a_i. Assumption (A3) may be less appropriate if cost of effort is a salient attribute in the setting being studied. For example, in Section 3.4 I use a quadratic cost of effort function because I am interested in how optimal aggregation rules change when the cost of effort function is strictly convex.

All that remains to make closed form solutions possible is to assign a functional form to the $\Pi(\vec{a})$, the expected future cash flows generated by the agent's actions. This specifi-

5. In section 3.5 I lift the background truncation assumption, as well as the square root utility assumption, and study a setting in which penalty schemes are allowed.

cation is best done within applied contexts, as the profit function can be tailored to fit the economics of the specific setting at hand. For example, tasks can be made complements in the profit function to represent benefits to job diversification, an approach that I take in section 3.3. The profit function can also be a useful tool for facilitating interior solutions, as I demonstrate in section 3.1.

2.3 Comparison to LEN

The LEN model imposes structure on the P-A problem by giving a specific form to the following elements: the distribution of signals, the utility function, the cost of effort function and the contract. My agency framework specifies each of these elements except for the contract. It is therefore LEN-*like* because it employs a parametric approach that provides the tractability needed to address interesting applied issues. However, this advantage is gained without LEN's potential drawback, that the contractual form imposed *ex ante* might not be truly optimal.

Table 1 gives a side-by-side comparison of my framework and the LEN framework. There are a few meaningful differences. First, contracts in the LEN model are (by assumption) linear in cash space, while contracts in my framework are (optimally) linear in utility space. Second, LEN employs normally-distributed performance measures while I use Poisson-distributed performance measures. I contend that the Poisson distribution is more appropriate than the normal for modeling accounting information. Both distributions allow for tractable aggregation, but the Poisson is bounded below by zero (as are accounts in a double-entry system), while the Normal is unbounded below. Finally, signal variance is exogenous in LEN models; the agent controls the signal's mean but no other moments of the distribution. In my framework, managerial actions influence all moments of the distribution; for example, any effect on an account's expected value renders an equal effect on the account's variance because $E(x) = Var(x) = \bar{x}$ for $x \sim Poisson(\bar{x})$. Allowing managerial effort to influence higher moments opens the door to more robust predictions as to how agency issues affect the distributions of earnings and other line items.

Table 1: Comparison of the LEN and Poisson frameworks.

	LEN framework	**My framework**
Distribution of performance measure x	$x \sim Normal(a, \sigma^2)$	$x \sim Poisson(\hat{x} + f(a))$
Contract $s(x)$	Linear *by assumption* in cash space	*Optimally* linear in utility space
Utility of consumption $u(s)$	Negative Exponential $u(s) = -e^{-s}$	Square root $u(s) = \sqrt{s}$
Cost of effort $c(a)$	Quadratic $c(a) = a^2$	Linear $c(a) = \delta a$
Composition of agent's utility $U(s, c(a))$	Multiplicatively separable $U(s, a) = -e^{-(s-a^2)}$	Additively separable $U(s, a) = \sqrt{s} - \delta a$

3

APPLICATIONS

This section provides a set of model applications to specific issues where accounting is at the heart. I focus on revenue and expense accounts - a reasonable starting point for representing more realistic accounting systems - and explore how the optimal use of these accounts as performance measures interacts with organizational design, firm strategy and aggregation rules. I assume Poisson-distributed accounts (A1) throughout all five applications. I use the tractability-oriented assumptions (A2 and A3) when they are appropriate for the setting being investigated.

Equation (2.10) shows that with assumptions (A1) and (A2), optimal contracts (in utility space) are linear in accounts; however, the optimal *weights* on those accounts depend on setting-specific assumptions. This allows for comparison of how accounts are optimally weighted across settings. Throughout the applications, I refer to an *optimal aggregation rule* as the weighting of accounts in the optimal contract. I define an aggregation rule as *neutral* if it weights revenues and expenses equally, *conservative* if it weights expenses more heavily, and *aggressive* if it weights revenues more heavily.[1]

3.1 Incentivizing growth and efficiency

What is the right balance between revenue growth and cost containment? This question is central to firm strategy and profitability, but to my knowledge it has not been addressed from a contracting perspective. I investigate it in the following simple setting: One manager is hired to increase expected revenues and decrease expected expenses. I denote these actions a_r and a_e and assume that they influence revenues and expenses, denoted x_r and x_e, through the following account technology.

$$x_r \sim Poisson(\bar{x}_r), \;\; \bar{x}_r = \hat{x}_r + a_r \tag{3.1}$$

$$x_e \sim Poisson(\bar{x}_e), \;\; \bar{x}_e = \hat{x}_e - a_e. \tag{3.2}$$

Absent managerial effort, expected revenues are $\hat{x}_r$; this exogenous component can be interpreted as expected revenues from existing customers or stable demand. Absent managerial effort, expected expenses are $\hat{x}_e$ and can include both productive spending and waste. Notice that a_e reduces expenses but does not impact revenues. This assumes that a_e does not involve cuts to productive spending, rather, it represents efforts made to improve operating efficiency, cut slack or renegotiate contracts with suppliers for better input prices. I assume that some positive amount of productive spending is required to maintain revenues and therefore x_e cannot be cut to zero.[2]

I assume that the profit function is

$$\Pi(a_r, a_e) = \pi a_r + \pi a_e + \pi_b a_r a_e, \tag{3.3}$$

1. There are factors exogenous to this model that may affect whether earnings should be conservative or aggressively constructed; when I refer to an aggregation rule as *optimal* it is exclusively from a contracting perspective.

2. In addition to being realistic, this assumption avoids the potential technical problem arising from the principal setting $a_e = \hat{x}_e$, which would allow the use of a forcing contract because $Var(x_e) = \hat{x}_e - a_e$ would equal zero.

where πa_r and πa_e respectively represent the expected future cash flows from a_r and a_e individually, and $\pi_b a_r a_e$ represents the additional expected future cash flows resulting from synergies between the two actions. With this profit function and assumptions (A1)-(A3), the principal's program is as follows.

$$\begin{aligned} &\max_{s(x_r,x_e),a_r,a_e} && \pi a_r + \pi a_e + \pi_b a_r a_e - E[s(x_r,x_e)|a_r,a_e] && \text{(OBJ)} \\ &\text{s.t} && E[\sqrt{s(x_r,x_e)}|a_r,a_e] - \delta_r a_r - \delta_e a_e \geq \bar{U} && \text{(IR)} \\ &\text{and} && \frac{\partial}{\partial a_r} E[\sqrt{s(x_r,x_e)}|a_r,a_e] = \delta_r && (\text{IC}_\text{r}) \\ &\text{and} && \frac{\partial}{\partial a_e} E[\sqrt{s(x_r,x_e)}|a_r,a_e] = \delta_e && (\text{IC}_\text{e}) \end{aligned} \tag{3.4}$$

The optimal contract for program (3.4) is provided in the following observation and follows directly from Lemma 1, with $\mu_i = 2\delta_i \bar{x}_i$ for $i = r, e$.

Observation 2. *The optimal contract solving program* (3.4) *is*

$$\sqrt{s} = \bar{U} + \delta_r a_r + \delta_e a_e + \delta_r (x_r - \bar{x}_r) + \delta_e (\bar{x}_e - x_e). \tag{3.5}$$

Notice that revenues and expenses are respectively weighted by δ_r and δ_e, the agent's marginal cost of effort on each task; intuitively, stronger incentives are required for tasks that are more personally costly to the agent.[3] If the agent is indifferent between actions, the optimal weighting rule is neutral. For simplicity, I assume that this is the case and let $\delta_r = \delta_e = \delta$ throughout the rest of the application.

If $\hat{x}_r$ and $\hat{x}_e$ are interpreted as prior period revenues and expenses, then expanding $\bar{x}_r$ and $\bar{x}_e$ and canceling some terms in contract (3.5) reveals that the optimal contract is linearly increasing in revenue growth and expense curtailment, or equivalently, is linear in positive earnings changes.

$$\sqrt{s} = \bar{U} + \delta \underbrace{(x_r - \hat{x}_r)}_{\substack{\text{increase in} \\ \text{revenues}}} + \delta \underbrace{(\hat{x}_e - x_e)}_{\substack{\text{decrease in} \\ \text{expenses}}} \tag{3.6}$$

3. This echoes the findings of Amershi, Banker, and Datar (1990), who show that the optimal way to aggregate signals for performance evaluation depends on the individual manager being evaluated.

Before solving for the second-best actions, it is worth observing that the first-best actions are equal: $a_r^* = a_e^*$. Absent contracting considerations, and with actions equally profitable in the function $\Pi = \pi(a_r) + \pi(a_e) + \pi_b a_r a_e$, the principal wants a perfectly even balance between revenue growth and cost cutting. The following lemma shows that agency costs upset this balance and cause the principal to set $a_e > a_r$.

Lemma 2. *The second-best actions solving program* (3.4) *are:*

$$a_r = \frac{\pi\pi_b - 2\pi_b\bar{U}\delta - \delta^2(4\delta^2 - \pi_b)}{\pi_b(4\delta^2 - \pi_b)} \tag{3.7}$$

$$\textit{and} \quad a_e = \frac{\pi\pi_b - 2\pi_b\bar{U}\delta + \delta^2(4\delta^2 - \pi_b)}{\pi_b(4\delta^2 - \pi_b)}. \tag{3.8}$$

The principal implements more cost cutting than revenue growth. The difference between the actions, given below, is decreasing in the synergy parameter (π_b) *and is increasing in the marginal cost of effort* (δ).

$$a_e - a_r = \frac{2\delta^2}{\pi_b} > 0. \tag{3.9}$$

Lemma 2 shows that, from a contracting perspective, cost cutting should be emphasized over revenue growth. This asymmetry stems from the opposite effect that the two actions have on the variances of their associated accounts and likelihood ratios. By equation (2.4), $Var(x_r) = \hat{x}_r + a_r$ and $Var(x_e) = \hat{x}_e - a_e$. Account variance is therefore increasing in a_r and decreasing in a_e. Let LR_r and LR_e be the likelihood ratios of the revenue and expense accounts, respectively. By equation (2.8), the variances of these likelihood ratios are as follows.

$$Var(LR_r) = \frac{1}{\bar{x}_r} = \frac{1}{\hat{x}_r + a_r} \tag{3.10}$$

$$Var(LR_e) = \frac{1}{\bar{x}_e} = \frac{1}{\hat{x}_e - a_e} \tag{3.11}$$

$Var(LR_r)$ is decreasing in the equilibrium revenue action (a_r), so as a_r increases, it becomes increasingly difficult to infer the agent's action from x_r. By contrast, $Var(LR_e)$ is increasing in the equilibrium cost-cutting action (a_e), making it easier to infer the agent's action from x_e. Notice that the finding that $a_e > a_r$ would not arise naturally from a LEN model because LEN assumes that variance is independent of managerial actions.

Because a_e is cheaper to implement than a_r, the existence of an interior solution relies on positive synergies between the actions. Setting $\pi_b = 0$ produces a corner solution, depicted in Figure 2. It shows that the principal's expected residual payoff is maximized by setting $a_r = 0$ and maximizing over a_e.With a linear profit function ($\pi a_r + \pi a_e$) and a linear cost of effort function, there is too little convexity for interior solutions when one action is cheaper to implement than the other.

The interior solution given by equations (3.7) and (3.8) obtains for any $\pi_b > 0$.[4] It seems reasonable to assume some nonzero benefit to balancing across action types and there is precedent for doing so; for example, Holmström and Milgrom (1991) assume that the principal wants the agent to put effort toward both quality and quantity.

Empirical insights: Focus on cost-cutting versus growth

The result from Lemma 2 that cost cutting will be emphasized over growth is supported by survey evidence in Graham, Harvey, and Rajgopal (2005). They document that when executives are asked what actions they would take when faced with an earnings target, 80 percent report that they would reduce spending while only 39 percent say they would try to increase revenues. My results shed light on the types of executives driving this result. Equation (3.9) shows that the emphasis on cost cutting becomes more extreme as marginal cost of effort increases. Interpreting an executive's marginal cost as an (inverse) expression of his or her ability, I predict that the least talented executives will focus almost entirely on cost cutting and put very little effort towards growth, while talented executives will take a more balanced approach towards the two objectives.

4. There are other ways to get interior solutions. For example, if the marginal product of the actions are different so that $\Pi = \pi_r a_r + \pi_e a_e$, interior solutions emerge when π_r is made sufficiently large relative to π_e. However, assuming a marginal positive benefit to balancing across actions seems more realistic and is mathematically simpler.

3.2 Revenue growth through sales volume versus selling price

Conceptualizing the revenue account as *price* $*$ *volume*, revenues can be increased through the *price component* or the *volume component.* In the account technology given in equations (3.1) and (3.2), an increase in revenues has no effect on expenses. This is akin to revenues being increased through the price component, as an increase in selling price does not affect production costs. I will refer to section 3.1 as the *price-focused setting* .

In this section I examine a *volume-focused setting.* Assume a manager is hired to take action a_e, which decreases expenses, and action a_v, which increases sales volume. Maintain all the assumptions of section 3.1, but replace equations (3.1) and (3.2) with the following account technology:

$$x_r \sim Poisson(\bar{x}_r), \quad \bar{x}_r = \hat{x}_r + a_v \tag{3.12}$$

$$\text{and} \quad x_e \sim Poisson(\bar{x}_e), \quad \bar{x}_e = \bar{x}_r\gamma - a_e = \hat{x}_e + \gamma a_v - a_e, \tag{3.13}$$

where $\gamma > 0$ is the exogenous cost margin, and consequently, exogenous expenses are $\hat{x}_e = \gamma\hat{x}_r$. Notice that a_v is reflected in both the x_r and x_e accounts; the manager improves revenues by increasing the number of units sold, and the added cost of producing those units appears in the expense account. The following observation shows that with this volume-increasing action, the optimal contract weights revenues more heavily than expenses.

Observation 3. *The optimal contract, shown below, uses an aggressive weighting rule, where the degree of aggression increasing in the ex ante cost margin.*

$$\sqrt{s} = \bar{U} + \delta(1+\gamma)(x_r - \hat{x}_r) + \delta(\hat{x}_e - x_e). \tag{3.14}$$

Revenues are optimally weighted by $\delta(1+\gamma)$ and expenses by δ. Because γ and δ are both positive, $\delta(1+\gamma) > \delta$ and therefore the optimal aggregation rule is always aggressive. This is in contrast with the neutral optimal aggregation rule in the price-focused setting (see contract (3.6)). Proper incentives for simultaneously motivating growth and cost containment therefore depend on the strategy used to achieve that growth.

The following lemma addresses the second-best balance between growth and cost-cutting in a volume-focused setting.

Lemma 3. *The principal implements more cost-cutting effort than volume-increasing effort; that is, $a_e > a_v$. The difference between the second-best actions (shown below) is increasing in the ex-ante cost margin γ.*

$$a_e - a_v = \frac{\delta^2 \left((1+\gamma) + (1+\gamma)^2\right)}{\pi_b} \tag{3.15}$$

As in the price-focused setting, the principal implements more expense reduction than revenue growth. Comparing equations (3.15) and (3.9) reveals that the emphasis on cost cutting is more pronounced in the volume-focused setting (as $\delta^2 \left((1+\gamma) + (1+\gamma)^2\right) > 2\delta^2$). This stronger emphasis on cost cutting stems from how expenses are influenced by a_v and a_e. Notice from equations (3.12) and (3.13) that a_v increases $\bar{x}_e$ while a_e decreases $\bar{x}_e$. Consequently, providing incentives for a_v indirectly produces incentives for a_e because it is in the agent's interest to combat the increased variance in the expense account arising from higher volume. These indirect incentives further reduce the agency costs associated with implementing a_e. Equation (3.15) reveals that the emphasis on cost-cutting is further exaggerated by higher cost margins (γ). This occurs because a larger cost margin enhances the influence of a_v on the expense account, thereby intensifying the indirect incentives for a_e.

Empirical predictions for price- versus volume-focused firms

Growth and cost containment are concerns likely shared by most firms. The findings in sections 3.1 and 3.2 suggest that, from an agency perspective, the proper way to balance and motivate growth and cost containment depends on firm strategy, specifically, whether firms are concerned with selling price or sales volume. I predict that volume-focused firms will use more aggressive earnings aggregates in their compensation contracts relative to price-focused firms. Furthermore, within samples of volume-focused firms, those with higher cost margins will weight revenues more aggressively in their compensation contracts, and the degree of aggression will increase with operating cost margin.

Most firms are unlikely to be exclusively price-focused or exclusively volume-focused but will fall on a continuum between the two extremes. Testing the predictions above requires some way to empirically estimate where firms fall on this spectrum. If we conceptualize price-focused firms as competing on product differentiation and volume-focused firms as competing on cost, then empirical strategies used to distinguish between differentiation-based and cost-based strategies in the competition literature could be applied here to identify volume- and price-focused firms. Conceptualizing price-focused firms as being more concerned with quality and volume-focused firms as more focused on quantity, textual analysis of management discussion and analysis in 10-K reports might help categorize firms accordingly. For example, price-focused firms might use more words and phrases like *product innovation, consumer experience, unique, loyalty*, or *luxury*, while volume-focused firms might use words and phrases like *efficient, lean, distribution channels, consumer value*, or *streamline.*

3.3 Job diversification and team size

How should tasks be grouped into jobs? And how should agents be grouped into teams? Holmström and Milgrom (1991) explored these questions in a simple two-agent, two-task setting and showed that agents should specialize in one task and never work in teams. Given that we observe teams in the real world, their model was clearly incomplete and they describe it as "merely a first pass" into studying optimal task grouping. Many papers since have continued to study the question of whether agents should specialize or work in teams. Most closely related to my analysis in this application is Hughes, Zhang, and Xie (2005). In a two-agent, two-task model, they study sufficient conditions for the principal to prefer diverse task assignment, in which both agents are assigned both tasks, as opposed to specific assignments, in which each agent specializes in one task. My research question diverges from papers like these by not asking *whether* agents should be assigned diverse tasks but rather *how diverse* their workloads should be. I examine the optimal team size when tasks are complements in the profit function and agents are evaluated by team output.

Consider a firm that produces and sells N products. Production takes place in a cost center that employs N homogeneous agents responsible for operating efficiency, and sales are managed by a revenue center that separately employs N homogeneous agents responsible for sales. Product sales are tracked in N separate revenue accounts, where x_{rj} denotes sales of product j. The expected value of x_{rj}, denoted $\bar{x}_{rj}$, is defined by the following technology, where a^r_{ij} is the revenue-increasing effort that agent i puts towards product j, and $\hat{x}_{rj}$ is the portion of expected revenues exogenous to the agents' actions.

$$\bar{x}_{rj} = \hat{x}_{rj} + \sum_{i=1}^{N} a^r_{ij}, \tag{3.16}$$

Similarly, the cost center tracks product costs in N individual expense accounts, where $\hat{x}_{ej}$ represents the inefficient level of production costs absent agent effort, and a^e_{ij} denotes the level of effort that agent i puts toward improving the operating efficiency of product j. Let x_{ej} denotes the operating costs of product j with the following expected value.

$$\bar{x}_{ej} = \hat{x}_{ej} - \sum_{i=1}^{N} a^e_{ij}, \tag{3.17}$$

I invoke assumption (A3) and assume that agents have no inherent preferences over projects; that is, projects are substitutes in the agents' (linear) cost functions. Letting δ be the marginal cost of effort across the homogeneous agents, agent i has cost of effort $c(\vec{a_i}) = \delta\,(a_{i1} + a_{i2} + ... + a_{iN})$.[5] I also invoke assumption (A2) and assume that all agents have square root utility preferences.

I first establish a benchmark case of task specialization in which each agent is assigned to only one product (and consequently has no team members). Assume that jobs are organized such that $a_{ik} = 0$ for all $k \neq i$; that is, agent i works only on product i. In this

5. I omit superscripts when an expression applies to both revenue and cost center actions/agents.

case, $\bar{x}_{ri} = \hat{x}_{ri} + a^r_{ii}$ and $\bar{x}_{ej} = \hat{x}_{ej} - a^e_{jj}$ for the jth and ith products in the revenue and cost centers, respectively. The principal contracts with each agent individually. By Lemma 1, the optimal contracts to elicit action a^r_{ii} from agent i in the revenue center and action a^e_{jj} from agent j in the cost center are as follows.[6]

$$\sqrt{s_i(x_{ri})} = \bar{U}_i + \delta a^r_{ii} + \delta(x_{ri} - \bar{x}_{ri}) \tag{3.18}$$

$$\sqrt{s_j(x_{ej})} = \bar{U}_j + \delta a^e_{jj} + \delta(\bar{x}_{ej} - x_{ej}) \tag{3.19}$$

Let the profit function of each product be linear in effort so that the principal's objective function when contracting with agent i is $\Pi_i = \pi a_i - E(s_i)$. Following the proof of Lemma 2, the second-best actions for the task specialization case are as follows.

$$a^{r*}_{ii} = \frac{\pi - 2\bar{U}\delta - \delta^2}{2\delta^2} \tag{3.20}$$

$$a^{e*}_{jj} = \frac{\pi - 2\bar{U}\delta + \delta^2}{2\delta^2} \tag{3.21}$$

Notice that because agents and products are homogeneous, actions are the same across agents within each center. Let $a_r \equiv a^{r*}_{ii}$ and $a_e \equiv a^{e*}_{jj}$ denote the second-best action for all agents in the revenue and cost centers, respectively.

Now assume that the principal is considering forming teams in order to take advantage of synergies, while maintaining the total effort level of a_r and a_e from each revenue and cost center agent. The principal would like to divide the revenue center into teams of m, where the members of each team work together on m products. Assume that teams are ordered such that agent 1 is assigned to products 1 through m and agent i is always assigned to product i. This team assignment process is shown below, where N agents are evenly divided into N/m teams.[7]

6. See the proof of Observation 2 for a straightforward derivation of this contract.

7. Assume that if N is not evenly divisible by m, the remaining agents are formed into team $N/m + 1$. For simplicity, I assume that the largest possible team size is $N/2$, and that N is large enough that the size of team $N/m + 1$ does not affect the principal's decision.

$$
\begin{array}{r}
\text{Agent} \\
\text{Product}
\end{array}
\underbrace{\begin{pmatrix} 1,2,...,m \\ 1,2,...,m \end{pmatrix}}_{\text{Team 1}}
\underbrace{\begin{pmatrix} m+1,...,2m \\ m+1,...,2m \end{pmatrix}}_{\text{Team 2}} ...
\underbrace{\begin{pmatrix} (k-1)m+1,...,km \\ (k-1)m+1,...,km \end{pmatrix}}_{\text{Team k}} ...
\underbrace{\begin{pmatrix} N-m+1,...,N \\ N-m+1,...,N \end{pmatrix}}_{\text{Team } N/m}
$$

Let team k be represented by the set $T_k \equiv \{(k-1)m+1,...,km\}$, where agent i is a member of team k if $i \in T_k$ and product j is assigned to team k if $j \in T_k$.

Assume that there are within-agent benefits of task diversification such that the principal's profit function for any agent i in team k is defined as follows, where the second term represents multiplicative synergistic benefit of diversifying agent i's task assignment.

$$\Pi_i = \sum_{j \in T_k} \pi a_{ij} + \pi_b \prod_{j \in T_k} a_{ij}, \ \forall i \in T_k. \tag{3.22}$$

The principal must choose m_r and m_e, the team size for the revenue and expense centers. The following lemma shows that $m_r < m_e$, where it is assumed that $\hat{x}_{rj} = \hat{x}_r$ and $\hat{x}_{ej} = \hat{x}_e$ for every product j.

Lemma 4. *Optimal team size is smaller for revenue tasks than for cost-cutting tasks; that is, $m_r < m_e$. For both task types, optimal team size is decreasing in the marginal cost of effort (δ) and in the account components exogenous to the agents' actions ($\hat{x}_r$ and $\hat{x}_e$).*

Equations (B.40) and (B.41) in the proof show that the expected net benefit of forming teams in the revenue and expense centers are as follows, where for ease of comparison I assume that $\hat{x}_r = \hat{x}_e \equiv \hat{x}$.

$$\text{Payoff}(m_r) = \pi_b \left(\frac{a_r}{m_r}\right)^{m_r} - \delta^2(m_r - 1)\left(\hat{x} + a_r\right) \tag{3.23}$$

$$\text{Payoff}(m_e) = \pi_b \left(\frac{a_e}{m_e}\right)^{m_e} - \delta^2(m_e - 1)\left(\hat{x} - a_e\right) \tag{3.24}$$

The first term in each equation is the benefit of job diversification. Recalling from equations (3.20) and (3.21) that $a_e > a_r$, the benefit of team formation is larger in the cost center than the revenue center. The second term in each equation is the cost of compensating each agent for the added risk they face when their contracts are written over $m-1$ additional products. Comparing the last terms in equations (3.23) and (3.24) reveals that the marginal compensation cost of increasing team size is smaller in the cost center than the revenue center. The result that $m_e > m_r$ therefore comes from cost center teams being both more beneficial and less costly than revenue center teams.

This application is not meant as comprehensive analysis of team formation and makes several simplifying assumptions, such as homogeneous agents and products, evenly divided teams, and identical account technology and action types within in each center.[8] Future work could relax these assumptions and address additional interesting questions. For example, in revenue versus cost centers, what agent characteristics are best suited for what types of products? Should agents with similar characteristics be grouped together, or is it beneficial to have diversity within teams? Future analyses could also address questions concerning team formation in profit centers; for example, whether it is better to form teams that specialize in cost-cutting versus sales or to form teams that do both types of actions and specialize by product.

Empirical predictions

The findings in section 3.3 predict that, on average, sales teams will be smaller than cost-cutting teams. A lack of data availability makes this prediction difficult to test, but it seems at least anecdotally consistent with reality. Consider a car manufacturer as an example. In Japanese automobile manufacturing practices such as the *Toyota Way*, every employee on the production floor is responsible for finding efficiency improvements; the entire production floor is essentially one large cost-cutting team. Car salesmen, by contrast, tend to work alone. A similar contrast might be drawn between door-to-door salespeople and the production teams that manufactured the goods they are selling. Given the obvious lack of archival data on this issue, field surveys might be required to determine whether sales

8. The assumption of homogeneous products is particularly objectionable; if products are perfectly homogeneous then there is really only one of them.

teams are in fact smaller than cost-cutting teams.

3.4 Aggregation rules and cost of effort functions

Lemma 1 reveals that optimal aggregation weights depend on two things: account technology and cost of effort. Section 3.2 experiments with account technology and shows how optimal aggregation rules changed as a result. In this application, I experiment with the agent's cost of effort; specifically, I abandon the weakly convex linear cost of effort function (A3) in favor of a strictly convex quadratic one.

Assume a manager is tasked with taking actions a_r and a_e to affect revenues and expenses via the account technology in equations (3.1) and (3.2). Let $c(a_r, a_e) = \frac{a_r^2}{2} + \frac{a_e^2}{2}$; notice that this quadratic cost of effort is additively separable in the two types of effort and is convex in each. Specify the profit function as $\Pi(a_r, a_e) = \pi a_r + \pi a_e$.[9] Maintaining the assumption of square root utility (A2), the optimal contract follows from Lemma 1 and is as presented in the following observation.

Observation 4. *The optimal contract, shown below, is a linear aggregation of accounts, where revenues are weighted by $c^r = a_r$ and expenses are weighted by $c^e = a_e$.*

$$\sqrt{s(x_r, x_e)} = \bar{U} + \frac{a_r^2}{2} + \frac{a_e^2}{2} + a_r(x_r - \bar{x}_r) + a_e(\bar{x}_e - x_e) \tag{3.25}$$

Interpreting $\hat{x}_r$ and $\hat{x}_e$ as last period's outcomes and rearranging equation (3.25), the contract is linear in the change in *weighted* earnings, where revenues are weighted by a_r and expenses by a_e.

$$\sqrt{s} = \bar{U} - \frac{a_r^2}{2} - \frac{a_e^2}{2} + \underbrace{\overbrace{(a_r x_r - a_e x_e)}^{\text{current weighted earnings}} - \overbrace{(a_r \hat{x}_r - a_e \hat{x}_e)}^{\text{prior weighted earnings}}}_{\text{change in weighted earnings}}$$

Let η represent the degree of conservatism in the aggregation rule, where the aggre-

9. The synergy term used in prior applications is no longer needed because the convexity provided by the quadratic cost of effort function is sufficient for avoiding corner solutions.

gation rule is conservative if $\eta > 1$, aggressive if $\eta < 1$ and neutral if $\eta = 1$. Because the contract weights revenues with a_r and expenses with a_e, deriving the optimal weights requires deriving the second-best actions. Substituting (3.25) into the principal's objective function gives an unconstrained maximization program, but unfortunately, with quadratic cost of effort the principal's program is too convex for me to derive closed-form expressions of the second-best actions. I am however able to obtain insight about the optimal weighting rule, as shown in the following lemma.

Lemma 5. *The degree of conservatism is characterized as follows:*

$$\eta = \frac{a_e}{a_r} = \frac{(a_r^2 + 3a_r + 2\hat{x}_r + a_e^2) + 2\bar{U}}{(a_e^2 - 3a_e + 2\hat{x}_e + a_r^2) + 2\bar{U}}. \tag{3.26}$$

The optimal aggregation rule is conservative unless $\hat{x}_e \gg \hat{x}_r$, and the degree of conservatism is increasing in $\hat{x}_r - \hat{x}_e$.

The lemma reveals that the optimal aggregation rule tends to be conservative because a_e tends to exceed a_r (for the reasons discussed in section 3.1). The degree of conservatism is increasing in $\hat{x}_r - \hat{x}_e$, the inherent profitability of the accounts under the manager's control.

An aggressive weighting rule may be optimal if the exogenous expenses are very large relative to exogenous revenues, i.e. $\hat{x}_e \gg \hat{x}_r$. Interpreting $\hat{x}_r$ and $\hat{x}_e$ as prior period outcomes, $\hat{x}_e \gg \hat{x}_r$ could represent the situation in which a CEO is hired to manage a firm that has been suffering extreme losses. Alternatively, $\hat{x}_e \gg \hat{x}_r$ could represent the situation in which a manager heads a sales department that is loaded with fixed costs (high $\hat{x}_e$) and operates in a competitive industry in which clients may be poached if their accounts are not regularly serviced (low $\hat{x}_r$).

Empirical predictions: Conservatism and Cost of Effort Convexity

Lemma 5 suggests that more profitable firms will compensate their CEOs on more conservative performance measures. To the extent that accounting policies are driven by contracting considerations, profitable firms will report more conservative earnings figures than loss-making firms.

Lemma 5 also shows that optimal aggregation rules tend to be conservative when cost of effort is quadratic; in contrast, the optimal aggregation rule is neutral when cost of effort is linear (see contract (3.6) in section 3.1). Imagine industries or occupations could be categorized by cost of effort convexity; for example, industries that are very competitive in the sense of product substitutability might be classified as having highly convex cost of effort. The contrast between sections 3.1 and 3.4 suggests that firms with more convex cost functions will use more conservative earnings aggregates in their compensation contracts. Similar cost of effort functions may help explain why firms in similar sectors tend to adopt similar performance measures, as documented by De Angelis and Grinstein (2015).

3.5 Penalty schemes

A result that emerged consistently in the previous applications is that it is cheaper to incentivize cost cutting than revenue growth. This application presents an extreme version of that finding by showing how cost cutting can be made approximately *free* from an agency cost perspective. I examine the cost-cutting and revenue growth settings separately. In each setting, I maintain the assumption of Poisson-distributed accounts (A1), but I abandon the tractability-oriented assumptions (A2) and (A3) in favor of more general cost of effort and utility representations. I assume that the cost of effort $c(\cdot)$ is increasing and convex and that the agent's utility function, $u(\cdot)$ is increasing and concave. Furthermore, I assume that the agent's utility function is unbounded below:

$$\lim_{s \to \underline{s}} u(s) = -\infty. \tag{3.27}$$

This is a relatively common assumption (e.g. Rogerson 1985, Assumption A.7) and is satisfied by well-known utility functions, such as the negative exponential used in LEN models.

I consider the cost-cutting setting first. Say that a risk-neutral principal hires a risk-averse agent to take action a_e with expense account technology $\bar{x}_e = \hat{x}_e - a_e$. I abandon the underlying assumption that accounts affected by mean-reducing actions are truncated at some arbitrarily high threshold; I let expenses be unbounded above so that $x_e \in \mathbb{N}$, the set of natural numbers.

Let (w^*, a_e^*) represent the first-best wage-action pair, where $w^* \equiv u^{-1}\left(\bar{U} + c(a_e^*)\right)$. The following lemma shows that the principal can approximate this first-best cost-cutting solution arbitrarily closely.

Lemma 6. *When a manager is hired to reduce waste, the first-best solution (w^*, a_e^*) can be implemented arbitrarily closely with a contract $(s(x_e), a_e)$ that takes the following form.*

$$s(x_e) = \begin{cases} \tilde{w} = w^* + \varepsilon & \text{if } x_e \leq M \\ K & \text{if } x_e > M \end{cases} \tag{3.28}$$
$$a_e = a_e^*$$

The contract stipulates that when operating expenses fall below threshold M, the agent is paid the first-best wage (w^*), plus some amount ε required to compensate the agent for the risk of incurring the penalty (thereby satisfying the IR constraint). For realizations above threshold M, the agent incurs penalty $K < w^*$. The proof shows that because the expense likelihood ratio is unbounded below, ε approaches zero as M approaches infinity. Therefore, because the probability of incurring the penalty disappears as M is taken to infinity, $E[\tilde{w}] = w^*$ in the limit. Assumption (3.27) is necessary for this result because it guarantees that for any threshold M, there is a penalty painful enough to make a_e^* incentive compatible. As the threshold increases, so does the brutality of K, as stated in the following corollary.

Corollary 1. *As the threshold M approaches infinity, the penalty K becomes infinitely harsh.*

A penalty scheme can also be used to enforce the first-best action when the agent's compensation is based on the aggregate signal, earnings, rather than on the operating expense account directly. Let earnings be represented by $y = x_r - x_e$, where $x_r = \hat{x}_r$ because revenues are not affected by the manager's action (a_e).

Lemma 7. *When a manager is hired to cut waste, the first-best solution (w^*, a_e^*) can be implemented arbitrarily closely with a contract $(s(y), a_e)$ that takes the following form.*

$$s(y) = \begin{cases} \tilde{w} = w^* + \delta & \text{if } y \geq y_M \\ K & \text{if } y < y_M \end{cases} \tag{3.29}$$

$$a_e = a_e^*$$

I now address the single-task setting in which the manager is hired to take action a_r to improve revenues through the baseline revenue technology $\bar{x}_r = \hat{x}_r + a_r$. Let (w^*, a_r^*) be the first-best wage-action pair. Approximating this first-best with a forcing contract is impossible.

Lemma 8. *The principal cannot approximate the first-best growth solution using a penalty scheme.*

Instead, the contract is monotonically increasing in revenues and mirrors the traditional second-best characterization of Holmstrom's (1979) equation (7).

Lemma 9. *If $(s(x_r), a_r)$ solves the principal's program, then $s(x_r)$ is increasing monotonically in x_r and satisfies the following equation for all x_r:*

$$\frac{1}{u'(s(x_r))} = \lambda_r + \mu_r \left(\frac{x_r - \bar{x}_r}{\bar{x}_r} \right). \tag{3.30}$$

There is a stark contrast between the optimal revenue contract (equation 3.30) and the expense contract (equation 3.28). The revenue contract is second-best: compensation is contingent on realized revenues and is inefficient from a risk-sharing perspective. The expense contract, by contrast, pays a flat wage in expectation and approximates the first-best solution. Cost-cutting actions are therefore cheaper (and almost free) to implement, and assuming that there is sufficient waste to cut, it is more profitable for the principal to hire a manager to cut waste than to grow revenues.

It is worthwhile to discuss why first-best approximation is possible for cost cutting but not growth. Notice that a_r is account-increasing while a_e is account-decreasing. This creates a difference in what it means to get a tail-end account balance when motivating a_e versus a_r. When motivating a_e, higher-than-expected operating expenses (x_e) is a bad outcome. As x_e approaches the extreme right tail, the principal becomes increasingly confident that the manager shirked, and in the limit she is *certain* that the agent shirked. In technical terms, the likelihood ratio approaches negative infinity as x_e approaches infinity. This near-certainty - in combination with the availability of penalties severe enough to deter shirking due to utility being unbounded below - allows the principal to approximate the first-best solution arbitrarily closely.

In contrast, when motivating growth actions, higher account balances are good; a bad outcome is one where revenues are less than expected. But there is a lower bound on bad revenues: the worst-case revenue outcome is zero dollars. When the principal observes lower-than-expected revenues, she is unsure whether the manager worked or shirked, even when revenues are zero. In technical terms, the revenue likelihood ratio is bounded below, and therefore the limiting argument made in Lemma 6 cannot be applied to the revenue case. The stark difference between the revenue and expense contracts is thus driven by the fact that accounts are bounded below by zero, and that a_r moves revenues away from that lower bound while a_e moves expenses closer to it.

Empirical predictions: Pay-performance sensitivities and life cycle

This section has shown that in certain conditions, cost-cutting actions are best implemented with flat wage penalty schemes, but that revenue growth is never efficiently implemented with these schemes. If one is willing to entertain the idea that at least *some* firms employ penalty schemes when implementing cost-cutting actions, then these findings suggest that agents charged with improving efficiency will have less variable compensation than those charged with growth. This leads to empirical predictions about pay-performance sensitivities (PPS) in settings where it is reasonable to expect that managers are charged primarily with either cost-cutting or revenues.

If Chief Operations Officers (COOs) are primarily tasked with efficiency and Chief Marketing Officers (CMOs) are primarily charged with revenue growth, then to the extent that at least some firms in a given sample employ penalty schemes to enact cost-cutting, the sensitivity of COO compensation to operating expenses should be lower than the sensitivity of CMO compensation to revenues. Notice that these predictions deal with sensitivities to *account*-level performance - expenses and revenues for COOs and CMOs, respectively. If CMOs have revenue-contingent contracts and COOs have expense-contingent contracts, then comparing performance sensitivities at the account-level should give sharper results than at the aggregate earnings level.

Life cycle plays an important role in the allocation of effort between cost cutting and revenue growth and therefore is relevant to predictions about PPS. Startup firms don't have waste to cut and are obviously concerned with growth. In pursuit of that growth, startups may require managers to make investments that are expected to pay off in the future but reduce current period earnings. It therefore seems unlikely that startups will contract on current accounting measures before revenues have substantially materialized.[10] This suggests that early startup firms will have low earnings-PPS. In fact, if managers are hired to make investments that flow through the income statement (e.g. research and development or marketing expenditures), it may actually appear empirically that the manager is being paid to generate losses. I predict that earnings-PPS is zero or even negative in samples of early startup firms and becomes less negative or turns positive as revenues materialize.

10. This idea is supported by evidence in De Angelis and Grinstein (2015), who find that firms with more growth opportunities rely on more market-based measures than accounting-based measures, and that mature firms are more likely to rely on accounting-based measures.

Lifecycle stage:	Startup/Growth	Maturity	Decline
Actions:	Revenue growth	Revenue growth Cost-cutting	Cost-cutting

Mature firms are likely concerned with both containing costs and increasing (or at least maintaining) revenues. If a sample of mature firms includes some firms that enforce cost-cutting through flat wage penalty schemes, CEO compensation in that sample should on average be more sensitive to revenues than to operating expenses.

As firms leave the maturity phase and seek to manage decline as profitably as possible, managers are likely to be tasked primarily with cost-cutting actions. Within samples declining firms, compensation should be more sensitive to costs than to revenues because revenue growth is less likely to be part of the manager's job description.

As firms that employ flat-wage penalty schemes to implement cost-cutting actions move from maturity to decline, compensation will continue to be flat in expenses and will stop varying in revenues. Therefore, as long as some firms employ flat-wage penalty schemes, earnings-PPS should decrease as firms move from maturity to decline. Suggestive evidence for this prediction can be found in Gilson and Vetsuypens (1993) who study financially distressed firms and find that earnings performance explains very little variation in CEO compensation. Similarly, Eckbo, Thorburn, and Wang (2016) find that CEOs retained during Ch 11 reorganization experience a median compensation change of zero. Carter, Hotchkiss, and Mohseni (2018) find that as firms become financially distressed they decrease their use of accounting-based metrics.

Empirical compensation studies have documented that CEO compensation is less sensitive to earnings when earnings are negative; that is, earnings PPS is lower in loss firms (e.g. Gaver and Gaver 1998; Leone, Wu, and Zimmerman 2006; Shaw and Zhang 2010). Several explanations have been put forth in the literature to explain this asymmetric sensitivity. Matějka, Merchant, and Van der Stede (2009) argue that managers of loss-making firms have shorter employment horizons and that consequently their compensation contracts are more likely to rely on nonfinancial performance measures rather than financial measures like earnings. Gaver and Gaver (1998) speculate that either CEOs at loss firms are extracting rents or compensation committees don't want to discourage loss-generating actions that are profitable in the long run. Drake, Engel, and Martin (2018) suggest that performance measures are less informative about managerial effort in loss years relative to profitable years.

My findings offer an alternative explanation: If more losses occur during startup and decline relative to maturity, then the lower PPS in loss firm-years relative to profitable firm-years can be explained in part by life cycle. There is nothing special about losses *per se*; if revenues and expenses are informative about a CEO's growth and cost-cutting efforts, this does not become less true in a year when earnings happen to be negative. Rather, it may just be that losses are more common during (1) the startup/growth stage when accounting measures are less likely to be used in compensation contracts, and (2) the decline phase, when managers are tasked primarily with cost-cutting actions which some firms might enforce using flat-wage penalty schemes.

What this section should make clear is that theory makes *no universal prediction* about the sensitivity of pay to performance in loss firms; I have discussed settings in which theory predicts earnings-level PPS to be negative, positive or zero. Therefore, treating loss firms as homogeneous by lumping them into one sample is unlikely to yield productive findings. A more fruitful approach might be to investigate the earnings components that are driving the losses.[11] This account-based perspective could shed light on what actions managers in a firm might be tasked with, as well as how optimal compensation is expected to vary with individual line items and with earnings.

11. Drake, Engel, and Martin (2018) take a different, clever approach in an attempt to differentiate among loss firms. They use deferred tax asset valuation allowances to categorize loss years by expected loss persistence.

4

CONCLUSION

Tasked with discussing the intellectual foundations of accounting research, John Fellingham identified two pervasive concepts in our discipline: *information* and *double entry mechanics*.[1] Analytical accounting research has focused on the former but neglected the latter. I believe this is an oversight. This paper demonstrates that the basic features of general ledger accounts should not be trivialized. Specifically, I show that accounts being bounded below by zero has contracting implications that affect task allocation, aggregation rules, team formation, and the relationship between pay and performance.

Future applications of my framework could investigate the organizational implications of managerial accounting practices, such as traditional versus activity-based costing or different joint cost allocation methods. With the exception of Lemma 7, this paper has focused exclusively on Poisson-distributed signals; future applications could make better use of the

1. Demski et al. 2002.

framework's aggregation technology by doing more with Skellam-distributed signals. I have considered an accounting system that perfectly reports true account realizations. Extensions of the model could incorporate a measurement element and show how accounting standards or reporting bias interact with broader organizational issues. For example, I have shown that revenue actions are more expensive to implement than cost-cutting actions; perhaps conservative accounting standards for revenue recognition could mitigate the inherent riskiness of revenue-increasing actions. The model could also be extended to incorporate a valuation or disclosure angle or to consider multiple time periods, perhaps where the realization of account x in period t becomes $\hat{x}$ in year $t+1$.

Finally, the applications in this paper have only considered income statement accounts. I hope in future work to better capture the dynamics of double-entry accounting rather than just the properties of accounts that reside in the double-entry system. By modeling the balance sheet alongside the income statement, I could perhaps shed light on how relationships among financial statement line items speak to the actions being taken by firm executives, providing a stewardship-oriented approach to financial statement analysis. In a multi-period extension of the framework, modeling both sides of each journal entry would allow me to study optimal accrual policies for incentivizing certain activities or providing information to different types of stakeholders.

APPENDIX A

THE SKELLAM DISTRIBUTION

A.1 Definition and properties

Let $y = x_1 - x_2$, where

$$x_1 \sim Poisson(\bar{x}_1), \quad \bar{x}_1 = \hat{x}_1 + f_1(\vec{a}) \tag{A.1}$$

$$\text{and} \quad x_2 \sim Poisson(\bar{x}_2), \quad \bar{x}_2 = \hat{x}_2 + f_2(\vec{a}). \tag{A.2}$$

Then y is said to follow the two-parameter Skellam distribution ($y \sim Skellam(\bar{x}_1, \bar{x}_2)$) and has the following probability mass function, which I call θ.

$$pr(y|\bar{x}_1, \bar{x}_2) = e^{-\bar{x}_1 - \bar{x}_2} \left(\frac{\bar{x}_1}{\bar{x}_2} \right)^{\frac{y}{2}} \mathrm{I}_y \left(2\sqrt{\bar{x}_1 \bar{x}_2} \right), \quad y = ..., -1, 0, 1, ... \tag{A.3}$$

where

$$\mathrm{I}_y(z) = \left(\frac{z}{2}\right)^y \sum_{j=0}^{\infty} \frac{\left(\frac{z^2}{4}\right)^j}{j!(y+j)!} \tag{A.4}$$

is the modified Bessel function of the first kind of order y. Following Alzaid and Omair (2010), I use the convention that any term containing a negative factorial in the denominator is equal to zero.

The Skellam distribution is unimodal. It is positively skewed when $\bar{x}_1 > \bar{x}_2$ and negatively skewed when $\bar{x}_1 < \bar{x}_2$. The first four moments of the $y \sim Skellam(\bar{x}_1, \bar{x}_2)$ are as follows.

$$E(y) = \bar{x}_1 - \bar{x}_2 \tag{A.5}$$

$$Var(y) = \bar{x}_1 + \bar{x}_2 \tag{A.6}$$

$$Skewness(y) = \frac{\bar{x}_1 - \bar{x}_2}{(\bar{x}_1 + \bar{x}_2)^{3/2}} \tag{A.7}$$

$$Kurtosis(y) = 3 + \frac{1}{\bar{x}_1 + \bar{x}_2} \tag{A.8}$$

A.2 Likelihood ratios for Skellam-distributed aggregates

Here I calculate the likelihood ratio of signal y with respect to action a_b, where y has the probability mass function shown in equation (A.3). I denote this likelihood ratio $LR_y^b \equiv \frac{\frac{\partial}{\partial a_b}\theta}{\theta}$. To find the derivative, I break equation (A.3) into three pieces and let $m = e^{-\bar{x}_1 - \bar{x}_2}$, $n = \left(\frac{\bar{x}_1}{\bar{x}_2}\right)^{k/2}$, and $o = I_k(2\sqrt{\bar{x}_1 \bar{x}_2})$; and I let m^b, n^b and o^b represent the derivatives of these terms with respect to a_b. Then

$$LR_y^b = \frac{m^b}{m} + \frac{n^b}{n} + \frac{o^b}{o}. \tag{A.9}$$

Now $m^b = (-1)\left[\frac{\partial}{\partial a_b}\bar{x}_1 + \frac{\partial}{\partial a_b}\bar{x}_2\right] e^{-\bar{x}_1 - \bar{x}_2}$. With $f_1^b = \frac{\partial}{\partial a_b}\bar{x}_1$ and $f_2^b = \frac{\partial}{\partial a_b}\bar{x}_2$, we have

$$\frac{m^b}{m} = -f_1^b - f_2^b \tag{A.10}$$

and

$$\frac{n^b}{n} = \frac{y}{2}\left(\frac{\bar{x}_2 f_1^b - \bar{x}_1 f_2^b}{\bar{x}_1 \bar{x}_2}\right). \tag{A.11}$$

The last term in equation (A.3), $o = I_y\left(2\sqrt{\bar{x}_1\bar{x}_2}\right)$ is the modified Bessel function of the first kind, which has the following property (Abramowitz and Stegun 1970, p. 376).

$$\frac{\partial}{\partial z} I_v(z) = I_{v+1}(z) + \frac{v}{z} I_v(z). \tag{A.12}$$

Then $o^b = \left(\frac{\partial}{\partial a_b}\left(2\sqrt{\bar{x}_1\bar{x}_2}\right)\right)\left[I_{y+1}(2\sqrt{\bar{x}_1\bar{x}_2}) + \frac{y}{2\sqrt{\bar{x}_1\bar{x}_2}} I_y(2\sqrt{\bar{x}_1\bar{x}_2})\right]$, and so

$$\frac{o^b}{o} = \frac{\bar{x}_1 f_2^b + \bar{x}_2 f_1^b}{(\bar{x}_1\bar{x}_2)^{1/2}}\left[\frac{I_{y+1}}{I_y} + \frac{y}{2\sqrt{\bar{x}_1\bar{x}_2}}\right] \tag{A.13}$$

Putting this all together gives

$$LR_y^b = \frac{\bar{x}_1 f_2^b + \bar{x}_2 f_1^b}{(\bar{x}_1\bar{x}_2)^{1/2}}\left[\frac{I_{y+1}}{I_y} + \frac{y}{2\sqrt{\bar{x}_1\bar{x}_2}}\right] + \frac{y}{2}\left(\frac{\bar{x}_2 f_1^b - \bar{x}_1 f_2^b}{\bar{x}_1\bar{x}_2}\right) - f_1^b - f_2^b. \tag{A.14}$$

APPENDIX B

PROOFS

Proof of Observation 1. Recall that $x_j \sim Poisson(\bar{x}_j)$; then

$$Pr(x_j|\bar{x}_j) = \frac{e^{-\bar{x}_j}(\bar{x}_j)^{x_j}}{x_j!}, \tag{B.1}$$

where $\bar{x}_j = \hat{x}_j + f_j(\vec{a})$. With $\frac{\partial \bar{x}_j}{\partial a_i} = \frac{\partial f_j(\vec{a})}{\partial a_i} \equiv f_j^i$, $LR_j^i \equiv \frac{\frac{\partial}{\partial a_i} Pr(x_j|\bar{x}_j)}{Pr(x_j|\bar{x}_j)}$ can be calculated as follows.

$$\frac{\frac{\partial}{\partial a_i}\left[Pr(x_j|\bar{x}_j)\right]}{Pr(x_j|\bar{x}_j)} = \frac{\frac{\partial}{\partial a_i}\left[(\bar{x}_j)^{x_j}\, e^{-\bar{x}_j}\right]}{(\bar{x}_j)^{x_j}\, e^{-\bar{x}_j}} = \frac{\frac{\partial}{\partial a_i}\left[(\bar{x}_j)^{x_j}\right] e^{-\bar{x}_j} + (\bar{x}_j)^{x_j}\, \frac{\partial}{\partial a_i}\left[e^{-\bar{x}_j}\right]}{(\bar{x}_j)^{x_j}\, e^{\bar{x}_j}}$$

$$= \frac{x_j f_j^i \left[(\bar{x}_j)^{x_j - 1}\right] e^{-\bar{x}_j} - (\bar{x}_j)^{x_j}\, f_j^i e^{-\bar{x}_j}}{(\bar{x}_j)^{x_j}\, e^{\bar{x}_j}}$$

$$= \frac{x_j f_j^i \left[(\bar{x}_j)^{x_j - 1} \right] - (\bar{x}_j)^{x_j} f_j^i}{(\bar{x}_j)^{x_j}}$$

$$= f_j^i \left[\frac{x_j (\bar{x}_j)^{x_j - 1} - (\bar{x}_j)^{x_j}}{(\bar{x}_j)^{x_j}} \right]$$

$$= f_j^i \left(\frac{x_j - \bar{x}_j}{\bar{x}_j} \right)$$

Now I prove the second part of the observation. By definition, $Var[LR] = E[LR^2] - [E(LR)]^2$. Noting that $[E(LR)]^2 = 0$, we have

$$\begin{aligned} Var[LR] = E[LR^2] &= E\left[\left(f_j^i \frac{x_j - \bar{x}_j}{\bar{x}_j} \right)^2 \right] \\ &= \frac{\left(f_j^i\right)^2}{\bar{x}_j^2} E\left[(x_j - \bar{x}_j)^2 \right] \\ &= \frac{\left(f_j^i\right)^2}{\bar{x}_j^2} E\left[x_j^2 - 2x_j\bar{x}_j + \bar{x}_j^2 \right]. \end{aligned}$$

By equations (2.3) and (2.4),

$$E[x_j^2] = \bar{x}_j + \bar{x}_j^2. \tag{B.2}$$

Using this to take the expectation,

$$\begin{aligned} Var[LR] &= \frac{\left(f_j^i\right)^2}{\bar{x}_j^2} \left(\bar{x}_j + \bar{x}_j^2 - 2\bar{x}_j^2 + \bar{x}_j^2 \right) \\ &= \frac{\left(f_j^i\right)^2}{\bar{x}_j^2} (\bar{x}_j) \\ &= \frac{\left(f_j^i\right)^2}{\bar{x}_j} \end{aligned} \tag{B.3}$$

$\square$

Proof of Lemma 1. Program (3.4) can be expressed in Lagrangian form as follows.

$$\begin{aligned}
\max_{s(\vec{x}),\vec{a}} \mathcal{L} =& \Pi(\vec{a}) - \sum_{x_1}\sum_{x_2}...\sum_{x_n} s(\vec{x}) \prod_{j=1}^{n} Pr(x_j|\vec{a}) \\
&+ \lambda \left[\sum_{x_1}\sum_{x_2}...\sum_{x_n} \sqrt{s(\vec{x})} \prod_{j=1}^{n} Pr(x_j|\vec{a}) - c(\vec{a}) - \bar{U} \right] \\
&+ \mu_1 \left[\sum_{x_1}\sum_{x_2}...\sum_{x_n} \sqrt{s(\vec{x})} \frac{\partial}{\partial a_1} \left(\prod_{j=1}^{n} Pr(x_j|\vec{a}) \right) - c^1 \right] \\
+...&+ \mu_i \left[\sum_{x_1}\sum_{x_2}...\sum_{x_n} \sqrt{s(\vec{x})} \frac{\partial}{\partial a_i} \left(\prod_{j=1}^{n} Pr(x_j|\vec{a}) \right) - c^i \right] \\
+...&+ \mu_m \left[\sum_{x_1}\sum_{x_2}...\sum_{x_n} \sqrt{s(\vec{x})} \frac{\partial}{\partial a_m} \left(\prod_{j=1}^{n} Pr(x_j|\vec{a}) \right) - c^m \right]
\end{aligned} \tag{B.4}$$

To simplify notation, let $P_j \equiv Pr(x_j|\vec{a})$, and let $P_j^i \equiv \frac{\partial}{\partial a_i} Pr(x_j|\vec{a})$. Notice that

$$\frac{\partial}{\partial a_i} (P_1 P_2 ... P_n) = P_1^i \prod_{j \neq i} P_j + P_2^i \prod_{j \neq 2} P_j ... + P_n^i \prod_{j \neq n} P_j.$$

Then differentiating (B.4) with respect to s and rearranging yields

$$\begin{aligned}
2\sqrt{s} = \lambda &+ \mu_1 \left(\frac{P_1^1}{P_1} + \frac{P_2^1}{P_2} + ... + \frac{P_n^1}{P_n} \right) + \mu_2 \left(\frac{P_1^2}{P_1} + \frac{P_2^2}{P_2} + ... + \frac{P_n^2}{P_n} \right) + ... \\
&... + \mu_m \left(\frac{P_1^m}{P_1} + \frac{P_2^m}{P_2} + ... + \frac{P_n^m}{P_n} \right).
\end{aligned}$$

Noting that, by definition, P_j^i / P_j is the likelihood ratio LR_j^i, and taking the calculation of LR_j^i from Observation 1, the above contract can be written as follows, which is the contract given in the lemma.

$$\begin{aligned}
2\sqrt{s} = \lambda &+ \mu_1 \sum_{j=1}^{n} f_j^1 \left(\frac{x_j - \bar{x}_j}{\bar{x}_j} \right) + \mu_2 \sum_{j=1}^{n} f_j^2 \left(\frac{x_j - \bar{x}_j}{\bar{x}_j} \right) + ... \\
&... + \mu_m \sum_{j=1}^{n} f_j^m \left(\frac{x_j - \bar{x}_j}{\bar{x}_j} \right).
\end{aligned} \tag{B.5}$$

I now solve for the multipliers $\mu_1, ..., \mu_m$. Substituting contract (B.5) into constraint (IC$_i$) gives

$$\frac{\partial}{\partial a_i}\left[\lambda + \mu_1 \sum_{j=1}^{n} f_j^1 \left(\frac{x_j - \bar{x}_j}{\bar{x}_j}\right) + \mu_2 \sum_{j=1}^{n} f_j^2 \left(\frac{x_j - \bar{x}_j}{\bar{x}_j}\right) + ... + \mu_m \sum_{j=1}^{n} f_j^m \left(\frac{x_j - \bar{x}_j}{\bar{x}_j}\right)\right] = 2c^i.$$

Noting that $\partial/\partial a_i E(x_j) = f_j^i$, carrying out the expectation in the expression above gives

$$\mu_1 \sum_{j=1}^{n} \left(\frac{f_j^1 f_j^i}{\bar{x}_j}\right) + \mu_2 \sum_{j=1}^{n} \left(\frac{f_j^2 f_j^i}{\bar{x}_j}\right) + ... + \mu_m \sum_{j=1}^{n} \left(\frac{f_j^m f_j^i}{\bar{x}_j}\right) = 2c^i, \quad \forall i = 1, ..., m$$

These m equations and m unknowns can be expressed in matrix form as $\mathrm{D}\boldsymbol{\mu} = \boldsymbol{c}'$, where $\boldsymbol{\mu} \equiv (\mu_1, ..., \mu_2)^{\mathrm{T}}$, $\boldsymbol{c}' \equiv (c^1 c^2 ... c^m)^{\mathrm{T}}$ and D is the coefficient matrix with $d_{ik} \equiv \frac{f_j^i f_j^k}{\bar{x}_j}$ as its the element in its ith row and kth column. Then the solutions given by equation (2.12) in the lemma follows directly from Cramer's method for solving systems of equations. $\square$

Proof of Observation 2. Program (3.4) can be expressed in Lagrangian form as follows.

$$\begin{aligned}\max_{s(x_r,x_e),a_r,a_e} \mathcal{L} = \quad & \pi a_r + \pi a_e + \pi_b a_r a_e - E[s(x_r,x_e)|a_r,a_e] \\ & +\lambda \left[E\left(\sqrt{s(x_r,x_e)}|a_r,a_e\right) - \delta_r a_r - \delta_e a_e - \bar{U}\right] \\ & +\mu_r \left[\tfrac{\partial}{\partial a_r} E[\sqrt{s(x_r,x_e)}|a_r,a_e] - \delta_r\right] \\ & +\mu_e \left[\tfrac{\partial}{\partial a_e} E[\sqrt{s(x_r,x_e)}|a_r,a_e] - \delta_e\right]\end{aligned} \tag{B.6}$$

Let $p = Pr(x_r|a_r)$ and $q = Pr(x_e|a_e)$; then because x_r and x_e are independent, $Pr(x_r \cap x_e|a_r,a_e) = pq$. Then taking the derivative of $\mathcal{L}$ with respect to s yields

$$2\sqrt{s} = \lambda + \mu_r \frac{p'}{p} + \mu_e \frac{q'}{q}. \tag{B.7}$$

By equation (2.7), $p'/p = \frac{x_r - \bar{x}_r}{\bar{x}_r}$ and $q'/q = (-1)\frac{x_e-\bar{x}_e}{\bar{x}_e} = \frac{\bar{x}_e - x_e}{\bar{x}_e}$; substituting these likelihood ratios into the contract above gives

$$2\sqrt{s} = \lambda + \mu_r \left(\frac{x_r - \bar{x}_r}{\bar{x}_r}\right) + \mu_e \left(\frac{\bar{x}_e - x_e}{\bar{x}_e}\right). \tag{B.8}$$

Now I solve for the multipliers. Notice that taking expectations over both sides of (B.8) gives

$$2E[\sqrt{s}] = \lambda. \tag{B.9}$$

The first-order condition of $\mathcal{L}$ with respect to λ gives

$$E\left(\sqrt{s}|a_r,a_e\right) - \delta_r a_r - \delta_e a_e = \bar{U} \tag{B.10}$$

Combining equations (B.9) and (B.10) gives $\lambda = 2\bar{U} + 2\delta_r a_r + 2\delta_e a_e$. Now substitute (B.8) into (IC_r) and (IC_e).

$$\frac{\partial}{\partial a_r} E\left[\frac{\lambda}{2} + \frac{\mu_r}{2}\left(\frac{x_r - \bar{x}_r}{\bar{x}_r}\right) + \frac{\mu_e}{2}\left(\frac{\bar{x}_e - x_e}{\bar{x}_e}\right)\right] = \delta_r \tag{B.11}$$

$$\frac{\partial}{\partial a_e} E\left[\frac{\lambda}{2} + \frac{\mu_r}{2}\left(\frac{x_r - \bar{x}_r}{\bar{x}_r}\right) + \frac{\mu_e}{2}\left(\frac{\bar{x}_e - x_e}{\bar{x}_e}\right)\right] = \delta_e \tag{B.12}$$

Because $E[x_r] = \hat{x}_r + a_r$ and $E[x_e] = \hat{x}_e - a_e$, we have $\frac{\partial}{\partial a_r} E[x_r] = 1$ and $\frac{\partial}{\partial a_e} E[x_e] = -1$. Then executing the expectations and partial derivatives in (B.11) and (B.12) and rearranging results in the following multipliers.

$$\mu_r = 2\delta_r \bar{x}_r \tag{B.13}$$

$$\mu_e = 2\delta_e \bar{x}_e \tag{B.14}$$

Finally, substituting λ, μ_r and μ_e into equation (B.8) gives the optimal contract presented in the lemma.

$\square$

Proof of Lemma 2. The optimal contract in utility space is given by equation (3.5). Squaring both sides gives the contract in cash space.

$$
\begin{aligned}
s = \left(\tfrac{\lambda}{2}\right)^2 \quad & +\delta_r^2(x_r-\bar{x}_r)^2+\delta_e^2(x_e-\bar{x}_e)^2+\lambda\delta_r(x_r-\bar{x}_r)+\lambda\delta_e(x_e-\bar{x}_e) \\
& +2\delta_r\delta_e(x_r-\bar{x}_r)(x_e-\bar{x}_e)
\end{aligned}
\tag{B.15}
$$

Now I expand the squared terms and take expectations over both sides.

$$
\begin{aligned}
E(s) = \left(\tfrac{\lambda}{2}\right)^2 \quad & +\delta_r^2\left(E[x_r^2]+\bar{x}_r^2-2\bar{x}_rE[x_r]\right)+\delta_e^2\left(E[x_e^2]+\bar{x}_e^2-2\bar{x}_eE[x_e]\right) \\
& +\lambda\delta_r\left(E[x_r]-\bar{x}_r\right)+\lambda\delta_e\left(x_e-\bar{x}_e\right) \\
& +2\delta_r\delta_e\left(E[x_r]-\bar{x}_r\right)\left(E[x_e]-\bar{x}_e\right)
\end{aligned}
\tag{B.16}
$$

We know that $E[x_r]=\bar{x}_r$ and $E[x_e]=\bar{x}_e$. Recalling that a Poisson distribution has the same variance as its mean, calculate $E(x_r^2)$ as follows.

$$
\begin{aligned}
& Var(x_r) && = E(x_r^2)-[E(x_r)]^2 \\
\iff \quad & \bar{x}_r && = E(x_r^2)-\bar{x}_r^2 \\
\iff \quad & E(x_r^2) && = \bar{x}_r+\bar{x}_r^2
\end{aligned}
\tag{B.17}
$$

Following the same procedure, $E(x_e^2)=\bar{x}_e+\bar{x}_e^2$. Then executing the expectations over x_r and x_e in (B.16) gives

$$E(s)=\left(\frac{\lambda}{2}\right)^2+\delta_r^2\bar{x}_r+\delta_e^2\bar{x}_e.$$

Finally, expanding $\bar{x}_r$, $\bar{x}_e$ and $\left(\frac{\lambda}{2}\right)^2$ gives

$$E(s)=\bar{U}^2+\delta_r^2a_r^2+\delta_e^2a_e^2+2\bar{U}\delta_ra_r+2\bar{U}\delta_ea_e+2\delta_r\delta_ea_ra_e+\delta_r^2(\hat{x}_r+a_r)+\delta_e^2(\hat{x}_e-a_e). \tag{B.18}$$

Substituting (B.18) into the principal's objective function leaves an unconstrained maximization program in which the principal chooses a_r and a_e to maximize the following expression.

$$\text{PP} = \pi(a_r + a_e) + \pi_b a_r a_e - \bar{U}^2 - \delta_r^2 a_r^2 - 2\bar{U}\delta_r a_r - \delta_e^2 a_e^2 - 2\bar{U}\delta_e a_e - 2\delta_r\delta_e a_r a_e$$
$$- \delta_r^2(\hat{x}_r + a_r) - \delta_e^2(\hat{x}_e - a_e)$$

Now I take first order conditions.

$$\left(\frac{\partial \text{PP}}{\partial a_r}\right): \quad \pi + \pi_b a_e - 2\delta_r^2 a_r - 2\bar{U}\delta_r - 2\delta_r\delta_e a_e - \delta_r^2 = 0 \tag{B.19}$$

$$\left(\frac{\partial \text{PP}}{\partial a_e}\right): \quad \pi + \pi_b a_r - 2\delta_e^2 a_e - 2\bar{U}\delta_e - 2\delta_r\delta_e a_r + \delta_e^2 = 0 \tag{B.20}$$

Solving this system of equations with $\delta_r = \delta_e = \delta$ gives the closed-form expressions provided in the lemma.

□

Proof of Observation 3. The principal's program is given as follows. I assume that $\delta_v = \delta_e = \delta$.

$$\begin{aligned}
\max_{s(x_r,x_e),a_v,a_e} \quad & \Pi(a_v,a_e) - E[s(x_r,x_e)|a_v,a_e] && \text{(OBJ)} \\
\text{subject to} \quad & E[\sqrt{s(x_r,x_e)}|a_v,a_e] - \delta a_v - \delta a_e \geq \bar{U} && \text{(IR)} \\
& \tfrac{\partial}{\partial a_v} E[\sqrt{s(x_r,x_e)}|a_v,a_e] = \delta && (\text{IC}_v) \\
& \tfrac{\partial}{\partial a_e} E[\sqrt{s(x_r,x_e)}|a_v,a_e] = \delta && (\text{IC}_e)
\end{aligned} \tag{B.21}$$

Let λ, μ_v and μ_e be the respective multipliers on constraints (IR), (IC$_v$) and (IC$_e$). Then differentiating the associated Lagrangian with respect to s gives the following, where LR_j^i denotes the likelihood ratio for account x_j with respect to action a_i.

$$2\sqrt{s} = \lambda + \mu_v \left(LR_r^v + LR_e^v\right) + \mu_e LR_e^e \tag{B.22}$$

By Observation 1, $LR_r^v = \frac{x_r - \bar{x}_r}{\bar{x}_r}$, $LR_e^v = \gamma\frac{x_e - \bar{x}_e}{\bar{x}_e}$ and $LR_e^e = \frac{\bar{x}_e - x_e}{\bar{x}_e}$.

$$2\sqrt{s} = \lambda + \mu_v \left[\frac{x_r - \bar{x}_r}{\bar{x}_r} + \gamma\left(\frac{x_e - \bar{x}_e}{\bar{x}_e}\right)\right] + \mu_e\left(\frac{\bar{x}_e - x_e}{\bar{x}_e}\right) \tag{B.23}$$

Now take expectations of (B.23) over x_r and x_e and substitute it into (IC$_v$), noting that $E(x_r) = \hat{x}_r + a_v$ and $E(x_e) = \gamma\hat{x}_r + \gamma a_v - a_e$.

$$\begin{aligned}
\frac{\partial}{\partial a_v}\Bigg[\lambda + \mu_v\left(\frac{\hat{x}_r + a_v - \bar{x}_r}{\bar{x}_r}\right) + \mu_v\gamma\left(\frac{(\gamma\hat{x}_r + \gamma a_v - a_e) - \bar{x}_e}{\bar{x}_e}\right) \\
+ \mu_e\left(\frac{\bar{x}_e - (\gamma\hat{x}_r + \gamma a_v - a_e)}{\bar{x}_e}\right)\Bigg] = 2\delta
\end{aligned}$$

Then taking the partial derivative on the left have side leaves

$$\mu_v\left(\frac{1}{\bar{x}_r} + \frac{\gamma^2}{\bar{x}_e}\right) - \mu_e\frac{\gamma}{\bar{x}_e} = 2\delta. \tag{B.24}$$

Following the same procedure for (IC_e) gives

$$\mu_v\left(\frac{-1}{\bar{x}_e}\right)+\mu_e\frac{1}{\bar{x}_e}=2\delta \tag{B.25}$$

Then (B.24) and (B.25) provide two equations and two unknowns; solving this system for μ_v and μ_e gives $\mu_v = 2\delta(1+\gamma)\bar{x}_r$ and $\mu_e = 2\delta\left[\bar{x}_e + \gamma(1+\gamma)\bar{x}_r\right]$.

Finally, substituting these multipliers into equation (B.23) and rearranging gives contract (3.14) in the observation.

$\square$

Proof of Lemma 3. The optimal contract in utility space is given by equation (3.14). Squaring both sides gives the contract in cash space.

$$\begin{aligned} s = \left(\tfrac{\lambda}{2}\right)^2 \quad &+\delta^2(1+\gamma)^2(x_r-\bar{x}_r)^2+\delta^2(x_e-\bar{x}_e)^2 \\ &+\lambda\delta(1+\gamma)(x_r-\bar{x}_r)+\lambda\delta(x_e-\bar{x}_e) \\ &+2\delta^2(1+\gamma)(x_r-\bar{x}_r)(x_e-\bar{x}_e) \end{aligned} \tag{B.26}$$

Now I expand the squared terms and take expectations over both sides.

$$\begin{aligned} E(s) = \left(\tfrac{\lambda}{2}\right)^2 \quad &+\delta^2(1+\gamma)^2\left(E[x_r^2]+\bar{x}_r^2-2\bar{x}_rE[x_r]\right) \\ &+\delta^2\left(E[x_e^2]+\bar{x}_e^2-2\bar{x}_eE[x_e]\right) \\ &+\lambda\delta(1+\gamma)\left(E[x_r]-\bar{x}_r\right)+\lambda\delta\left(x_e-\bar{x}_e\right) \\ &+2\delta^2(1+\gamma)\left(E[x_r]-\bar{x}_r\right)\left(E[x_e]-\bar{x}_e\right) \end{aligned} \tag{B.27}$$

Following the logic shown in equation (B.17) from the proof of Lemma 2, I execute the expectations and I'm left with:

$$E(s) = \left(\frac{\lambda}{2}\right)^2 + \delta^2(1+\gamma)^2\bar{x}_r + \delta^2\bar{x}_e.$$

Finally, expanding $\bar{x}_r$, $\bar{x}_e$ and $\left(\frac{\lambda}{2}\right)^2$ gives

$$\begin{aligned} E(s) \quad &= \bar{U}^2+\delta^2a_v^2+\delta^2a_e^2+2\bar{U}\delta a_v+2\bar{U}\delta a_e+2\delta^2a_va_e \\ &+\delta^2(1+\gamma)^2(\hat{x}_r+a_v)+\delta^2(\gamma\hat{x}_r+\gamma a_v-a_e). \end{aligned} \tag{B.28}$$

Substituting (B.28) into the principal's objective function leaves an unconstrained maximization program in which the principal chooses a_v and a_e to maximize the following expression.

$$\begin{aligned} \mathrm{PP} = \pi(a_v+a_e)+\pi_ba_va_e-\bar{U}^2-\delta^2a_v^2-2\bar{U}\delta a_v-\delta^2a_e^2-2\bar{U}\delta a_e-2\delta^2a_va_e \\ -\delta^2(1+\gamma)^2(\hat{x}_r+a_v)-\delta^2(\gamma\hat{x}_r+\gamma a_v-a_e) \end{aligned}$$

Now I take first order conditions.

$$\left(\frac{\partial \mathrm{PP}}{\partial a_v}\right): \quad \pi + \pi_b a_e - 2\delta^2 a_v - 2\bar{U}\delta - 2\delta^2 a_e - \delta^2(1+\gamma)^2 - \delta^2\gamma = 0 \tag{B.29}$$

$$\left(\frac{\partial \mathrm{PP}}{\partial a_e}\right): \quad \pi + \pi_b a_v - 2\delta^2 a_e - 2\bar{U}\delta - 2\delta^2 a_v + \delta^2 = 0 \tag{B.30}$$

Solving this system of equations gives the closed-form expressions for a_v and a_e provided in the lemma.

$\square$

Proof of Lemma 4. Recall that the principal wants to incentivize total effort a_r and a_e from each agent in the revenue and expense centers, respectively. With homogeneous agents and homogeneous products, the principal can do no better than having each agent within a team split her effort evenly among the m products assigned to that team.

$$a_{ij}^r = \begin{cases} \frac{a_r}{m_r} & \text{if } i,j \in T_k \\ 0 & \text{otherwise} \end{cases} \tag{B.31}$$

$$a_{ij}^e = \begin{cases} \frac{a_e}{m_e} & \text{if } i,j \in T_k \\ 0 & \text{otherwise} \end{cases} \tag{B.32}$$

With this action symmetry and the homogeneity of agents and products, I simplify the analysis by focusing on the principal's contracting problem with agent 1. I rewrite (3.18) to obtain the optimal contract for agent 1 in the revenue center.

$$\sqrt{s_1} = \bar{U} + \delta a_r + \delta \sum_{j=1}^{m} (x_{rj} - \bar{x}_{rj}). \tag{B.33}$$

Squaring both sides of equation (B.33) gives the contract in cash space.

$$s_1 = \bar{U}^2 + 2\bar{U}\delta a_r + \delta^2 a_r^2 + \delta^2 \sum_{j=1}^{m} (x_{rj} - \bar{x}_{rj})^2 + \left(2a_r\delta^2 + 2\bar{U}\delta\right) \sum_{j=1}^{m} (x_{rj} - \bar{x}_{rj}) \tag{B.34}$$

Taking expectations over x_{rj} gives $E[s_1]$.

$$\begin{aligned}
& E(s_1) = \bar{U}^2 + 2\bar{U}\delta a_r + \delta^2 a_r^2 + \delta^2 \sum_{j=1}^{m_r} \left(E[x_{rj}^2] + \bar{x}_{rj}^2 - E[x_{rj}]\bar{x}_{rj} \right) \\
\iff \; & E(s_1) = \bar{U}^2 + 2\bar{U}\delta a_r + \delta^2 a_r^2 + \delta^2 \sum_{j=1}^{m_r} \left(E[x_{rj}^2] + \bar{x}_{rj}^2 - 2E[x_{rj}]\bar{x}_{rj} \right) \\
\iff \; & E(s_1) = \bar{U}^2 + 2\bar{U}\delta a_r + \delta^2 a_r^2 + \delta^2 \sum_{j=1}^{m_r} \left(\bar{x}_{rj} + \bar{x}_{rj}^2 + \bar{x}_{rj}^2 - 2\bar{x}_{rj}^2 \right) \\
\iff \; & E(s_1) = \bar{U}^2 + 2\bar{U}\delta a_r + \delta^2 a_r^2 + \delta^2 \sum_{j=1}^{m_r} \bar{x}_{rj} \qquad \text{(B.35)}
\end{aligned}$$

When analyzing the principal's contracting program with a focal agent (here agent 1), I denote the action of any other agent as $\hat{a}_{ij}$. Then the expected revenue of product $j \in T_1$ can be expressed as

$$\begin{aligned}
\bar{x}_{rj} &= \hat{x}_{rj} + \sum_{i=2}^{m_r} \hat{a}_{ij}^r + \frac{a_r}{m_r} && \text{(B.36)} \\
&= \hat{x}_{rj} + \sum_{i=2}^{m_r} \frac{\hat{a}_r}{m_r} + \frac{a_r}{m_r} && \text{(B.37)} \\
&= \hat{x}_r + (m_r - 1)\frac{\hat{a}_r}{m_r} + \frac{a_r}{m_r}, && \text{(B.38)}
\end{aligned}$$

where the second equality comes from equation (B.31), and the third equality comes from homogeneous products with $\hat{x}_{rj} = \hat{x}_r$ for all j. Substituting (B.38) into equation (B.35) and substituting action (B.31) into the profit function (3.22) gives PP_1^r, the principal's expected payoff from contracting with agent 1 in the revenue center.

$$PP_1^r = \pi a_r + \pi_b \left(\frac{a_r}{m_r} \right)^{m_r} - \bar{U}^2 - 2\bar{U}\delta a_r - \delta^2 a_r^2 - \delta^2 \sum_{j=1}^{m_r} \left(\hat{x}_r + (m_r - 1)\frac{\hat{a}_r}{m_r} + \frac{a_r}{m_r} \right)$$

Homogeneity allows the following simplification.

$$PP_1^r = \pi a_r + \pi_b \left(\frac{a_r}{m_r} \right)^{m_r} - \bar{U}^2 - 2\bar{U}\delta a_r - \delta^2 a_r^2 - \delta^2 \left[m_r \hat{x}_r + (m_r - 1)\hat{a}_r + a_r \right] \qquad \text{(B.39)}$$

Equation (B.39) shows that in determining team size, the principal faces a tradeoff between the synergistic benefit of job diversification, term $\pi_b (a_r/m_r)^{m_r}$, and the added cost of compensating the agent for team output, which is $\delta^2(m_r - 1)(\hat{x}_r + \hat{a}_r)$. Compensating for team output adds risk to the agent's contract because her compensation is now dependent on $m_r - 1$ additional products, each with exogenous variance $\hat{x}_r$ and actions $\hat{a}_r$ that are out of the agent's control. With homogeneous agents and evenly divisible teams, $PP_1^r = PP_i^r$ for any agent i in the revenue center. Analysis of the expected payoff from a single agent is therefore equivalent to analysis of cumulative payoff from contracting with all N agents, and with $\hat{a}_r = a_r$, the principal's expected payoff can be rewritten as

$$PP_r = \pi a_r + \pi_b \left(\frac{a_r}{m_r}\right)^{m_r} - \bar{U}^2 - 2\bar{U}\delta a_r - \delta^2 a_r^2 - \delta^2 m_r (\hat{x}_r + a_r) \tag{B.40}$$

Following the same steps for the cost center gives the principal's expected payoff from contracting with agents in the expense center.

$$PP_e = \pi a_e + \pi_b \left(\frac{a_e}{m_e}\right)^{m_e} - \bar{U}^2 - 2\bar{U}\delta a_e - \delta^2 a_e^2 - \delta^2 m_e (\hat{x}_e - a_e) \tag{B.41}$$

Comparing equations (B.40) and (B.41) reveals that the marginal cost of increasing team size is $\delta^2 (\hat{x}_r + a_r)$ for the revenue center and $\delta^2 (\hat{x}_e - a_e)$ for the expense center. Assume for simplicity that $\hat{x}_r = \hat{x}_e = \hat{x}$, and recall that $a_e > a_r$ (see equations (3.20) and (3.21)). Then the benefits of teamwork are larger for the cost center because $\pi_b \left(\frac{a_e}{m}\right)^m > \pi_b \left(\frac{a_r}{m}\right)^m$, and the costs of teamwork are smaller for the cost center because $\delta^2 m(\hat{x} - a_e) < \delta^2 m(\hat{x} + a_r)$. Teams are therefore larger in the cost center than in the revenue center, i.e. $m_e > m_r$. Additionally, teams will be smaller when $\hat{x}$ or δ is large because the cost of team size is increasing in $\hat{x}$ and δ.

□

Proof of Observation 4. The principal's program can be written as follows.

$$\begin{array}{rll}
\max\limits_{s(x_r,x_e),a_r,a_e} & E[\Pi(a_r,a_e)] - E[s|a_r,a_e] & \text{(OBJ)} \\
\text{s.t} & E[\sqrt{s}|a_r,a_e] - \frac{a_r^2}{2} - \frac{a_r^2}{2} \geq \bar{U} & \text{(IR)} \\
\text{and} & \frac{\partial}{\partial a_r} E[\sqrt{s}|a_r,a_e] = a_r & (\mathrm{IC_r}) \\
\text{and} & \frac{\partial}{\partial a_e} E[\sqrt{s}|a_r,a_e] = a_e & (\mathrm{IC_e})
\end{array}$$

Let λ, μ_r and μ_e be the multipliers on the IR, $\mathrm{IC_r}$ and $\mathrm{IC_e}$ constraints, respectively. Then differentiating the Lagrangian pointwise with respect to s yields

$$2\sqrt{s} = \lambda + \mu_r \left(\frac{x_r - \bar{x}_r}{\bar{x}_r} \right) + \mu_e \left(\frac{\bar{x}_e - x_e}{\bar{x}_e} \right), \tag{B.42}$$

where the terms in parentheses are the likelihood ratios of the revenue and expense accounts. Taking the expectation over both sides of (B.42) and noticing that the expected values of the likelihood ratios equal zero, we get $2E[\sqrt{s}] = \lambda$. Substituting this into the (binding) IR constraint yields

$$\lambda = 2\bar{U} + a_r^2 + a_e^2. \tag{B.43}$$

Substituting (B.42) into ($\mathrm{IC_r}$) gives

$$\begin{aligned}
& \mu_r \tfrac{\partial}{\partial a_r} \left(E\left[\tfrac{x_r}{\bar{x}_r} \right] \right) = 2a_r \\
\Longleftrightarrow \quad & \mu_r \tfrac{\partial}{\partial a_r} \left(\tfrac{\hat{x}_r + a_r}{\bar{x}_r} \right) = 2a_r \\
\Longleftrightarrow \quad & \tfrac{\mu_r}{\bar{x}_r} = 2a_r \\
\Longleftrightarrow \quad & \mu_r = 2a_r \bar{x}_r.
\end{aligned} \tag{B.44}$$

Similarly, substituting (B.42) into ($\mathrm{IC_e}$) gives $\mu_e = 2a_e\bar{x}_e$. Now substituting these expressions for λ, μ_r and μ_e back into (B.42), dividing through by 2 and rearranging gives

$$\sqrt{s} = \bar{U} + \frac{a_r^2}{2} + \frac{a_r^2}{2} + a_r(x_r - \bar{x}_r) + a_e(\bar{x}_e - x_e). \tag{B.45}$$

Finally, substituting in $\bar{x}_r = \hat{x}_r + a_r$ and $\bar{x}_e = \hat{x}_e - a_e$ and rearranging yields equation (3.25) in the observation. $\square$

Proof of Lemma 5.

The principal wants to maximize $E[\Pi|a_r, a_e] - E[s|a_r, a_e]$ over a_r and a_e. To find $E[s|a_r, a_e]$, first square (B.45) and take the expectation over x_r and x_e.

$$E[s] = \tfrac{\lambda^2}{4} + a_r^2 E(x_r^2) + a_r^2 \bar{x}_r^2 - 2a_r^2 \bar{x}_r E(x_r) + a_e^2 \bar{x}_e^2 + a_e^2 E(x_e^2) - 2a_e^2 \bar{x}_e E(x_e) \tag{B.46}$$

By equation (B.17), $E(x_r^2) = \bar{x}_r + \bar{x}_r^2$ and $E(x_e^2) = \bar{x}_e + \bar{x}_e^2$. Then (B.46) can be rewritten as follows.

$$\begin{aligned} & E[s] = \tfrac{\lambda^2}{4} + a_r^2 \left(\bar{x}_r + \bar{x}_r^2\right) + a_r^2 \bar{x}_r^2 - 2a_r^2 \bar{x}_r^2 + a_e^2 \bar{x}_e^2 + a_e^2 \left(\bar{x}_e + \bar{x}_e^2\right) - 2a_e^2 \bar{x}_e^2 \\ \iff \quad & E(s) = \tfrac{\lambda^2}{4} + a_r^2 \bar{x}_r + a_e^2 \bar{x}_e \end{aligned}$$

Now substituting in λ, $\bar{x}_r$ and $\bar{x}_e$ gives

$$E(s) = \bar{U}^2 + \tfrac{a_r^4}{4} + \tfrac{a_e^4}{4} + \bar{U}a_r^2 + \bar{U}a_e^2 + \tfrac{a_r^2 a_e^2}{2} + a_r^2 \left(\hat{x}_r + a_r\right) + a_e^2 \left(\hat{x}_e - a_e\right)$$

and the principal chooses a_r and a_e to maximize her expected payoff, given in the following expression.

$$\pi(a_r + a_e) - \bar{U}^2 - \tfrac{a_r^4}{4} - \tfrac{a_e^4}{4} - \bar{U}a_r^2 - \bar{U}a_e^2 - \tfrac{a_r^2 a_e^2}{2} - a_r^2 \hat{x}_r - a_r^3 - a_e^2 \hat{x}_e + a_e^3. \tag{PP}$$

Differentiating with respect to a_r and a_e yields the following first-order conditions.

$$\left(\frac{\partial \mathrm{PP}}{\partial a_r}\right): \quad \pi = a_r^3 + 2\bar{U}a_r + a_e^2 a_r + 2\hat{x}_r a_r + 3a_r^2 \tag{B.47}$$

$$\left(\frac{\partial \mathrm{PP}}{\partial a_e}\right): \quad \pi = a_e^3 + 2\bar{U}a_e + a_r^2 a_e + 2\hat{x}_e a_e - 3a_e^2 \tag{B.48}$$

The two equations appear symmetric except for the minus sign before the last term. Assume for a moment that $\hat{x}_r = \hat{x}_e$; then it cannot be the case that $a_r = a_e$ after cancellations we'd be left with $3a_r^2 = -3a_e^2$, whose only solution is the degenerate $a_r = a_e = 0$. With $a_r \neq a_e$, the only way for the equations to hold is to have $a_e > a_r$ so that the positivity of a_e^3 is large enough to overwhelm the negativity of $-3a_e^2$.

The one thing that can disrupt the finding that $a_e > a_r$ is to have $\hat{x}_e \gg \hat{x}_e$; if $\hat{x}_e$ becomes very large relative to $\hat{x}_r$ there reaches a point which requires $a_r > a_e$ for the equations to hold.

Setting the right-hand sides of (B.47) and (B.48) equal and rearranging provides the characterization of a_e/a_r shown in Lemma 5.

$\square$

Proof of Lemma 6. To implement the contract in Lemma 6, the payments $\tilde{w}$ and K must ensure that the agent is willing to accept the contract and the contract is incentive compatible with action a_e^*. That is, K and $\tilde{w}$ must satisfy the following two conditions, given threshold M, where I let $q_j(a_e) \equiv Pr(x_e = j|a_e)$ and $q_j^e(a_e) \equiv \frac{\partial}{\partial a_e} Pr(x_e = j|a_e)$

$$\sum_{j=0}^{M} u(\tilde{w}) q_j(a_e^*) + \sum_{j=M+1}^{\infty} u(K) q_j(a_e^*) = u(w^*) \tag{B.49}$$

$$\sum_{j=0}^{M} u(\tilde{w}) q_j^e(a_e^*) + \sum_{j=M+1}^{\infty} u(K) q_j^e(a_e^*) = c'(a_e^*) \tag{B.50}$$

Let $Q_i \equiv \sum_{j=0}^{i} q_j$ and $Q_i^e \equiv \sum_{j=0}^{i} q_j^e$ represent the expense account CDF and the derivative of the CDF with respect to a_e. By the definition of $q_j(a_e)$ as a probability density function, $\sum_{j=0}^{\infty} q_j^e(a_e) = 0$; this implies that $\sum_{j=M+1}^{\infty} q_j^e(a_e^*) = -\sum_{j=0}^{M} q_j^e(a_e^*) = -Q_M^e$. Then solving equation (B.50) for $u(K)$ and substituting into (B.49) yields:

$$u(\tilde{w}) = u(w^*) + c'(a_e^*) \frac{(1 - Q_M)}{Q_M^e} \tag{B.51}$$

$$u(K) = u(w^*) - c'(a_e^*) \frac{Q_M}{Q_M^e} \tag{B.52}$$

Therefore, the contractual payments are characterized as follows.

$$\tilde{w} \equiv w^* + \varepsilon = u^{-1} \left(u(w^*) + c'(a_e^*) \frac{1 - Q_M}{Q_M^e} \right). \tag{B.53}$$

$$K = u^{-1} \left(u(w^*) - c'(a_e^*) \frac{Q_M}{Q_M^e} \right). \tag{B.54}$$

The first-best solution is the pair (w^*, a_e^*) which is what the penalty scheme is aimed to approximate. Conditions (B.49) and (B.50) ensure that the agent will take action a_e^*; the remainder of the proof is dedicated to showing that the payment $\tilde{w}$ can be made arbitrarily close to w^* by increasing the penalty threshold. I will establish that $\tilde{w}$ approaches w^* from above as M approaches infinity, i.e. $\varepsilon \longrightarrow 0$ as $M \longrightarrow \infty$. Equation (B.53) shows that this limit obtains if $c'(a_e^*)\frac{1-Q_M}{Q_M^e}$ approaches zero. With $c'(a_e^*)$ constant in M, what remains to show is that $\frac{1-Q_M}{Q_M^e}$ approaches zero as M goes to infinity. The following calculation shows that that $Q_i^e = q_i$.

$$
\begin{aligned}
Q_i^e &\equiv \textstyle\sum_{j=0}^{i} q_j^e \\
&= \textstyle\sum_{j=0}^{i} \frac{e^{-(\bar{x}_e)}(\bar{x}_e)^j}{j!}\left[1-\frac{j}{(\bar{x}_e)}\right] \\
&= e^{-(\bar{x}_e)}\left[\frac{(\bar{x}_e)^0}{0!}\left(1-\frac{0}{(\bar{x}_e)}\right)+\frac{(\bar{x}_e)^1}{1!}\left(1-\frac{1}{(\bar{x}_e)}\right)+\frac{(\bar{x}_e)^2}{2!}\left(1-\frac{2}{(\bar{x}_e)}\right)+\ldots+\frac{(\bar{x}_e)^i}{i!}\left(1-\frac{i}{(\bar{x}_e)}\right)\right] \\
&= e^{-(\bar{x}_e)}\left[1+(\bar{x}_e-1)+\left(\frac{(\bar{x}_e)^2}{2!}-\bar{x}_e\right)+\left(\frac{(\bar{x}_e)^3}{3!}-\frac{(\bar{x}_e)^2}{2!}\right)+\ldots+\left(\frac{(\bar{x}_e)^i}{i!}-\frac{(\bar{x}_e)^{i-1}}{(i-1)!}\right)\right] \\
&= e^{-(\bar{x}_e)}\left[\cancel{1}+\cancel{(\bar{x}_e)}-\cancel{1}+\cancel{\frac{(\bar{x}_e)^2}{2!}}-\cancel{(\bar{x}_e)}+\cancel{\frac{(\bar{x}_e)^3}{3!}}-\cancel{\frac{(\bar{x}_e)^2}{2!}}+\ldots+\frac{(\bar{x}_e)^i}{i!}-\cancel{\frac{(\bar{x}_e)^{i-1}}{(i-1)!}}\right] \\
&= \frac{e^{-(\bar{x}_e)}(\bar{x}_e)^i}{i!} \\
&= q_i
\end{aligned}
$$

Then $\frac{1-Q_M}{Q_M^e}$ can be rearranged as follows.

$$
\begin{aligned}
\frac{1-Q_M}{Q_M^e} &= \frac{1-\sum_{j=0}^{M} q_j}{q_M} \\
&= \frac{1-\sum_{j=0}^{M}\frac{(\bar{x}_e)^j e^{-\bar{x}_e}}{j!}}{\frac{(\bar{x}_e)^M e^{-\bar{x}_e}}{M!}} \\
&= \frac{M!}{(\bar{x}_e)^M e^{-\bar{x}_e}}\left(1-e^{-\bar{x}_e}\textstyle\sum_{j=0}^{M}\frac{(\bar{x}_e)^j}{j!}\right) \\
&= \frac{M!}{(\bar{x}_e)^M}\left(e^{\bar{x}_e}-\textstyle\sum_{j=0}^{M}\frac{(\bar{x}_e)^j}{j!}\right) \\
\implies \quad & \lim_{M\to\infty}\left[\frac{M!}{(\bar{x}_e)^M}\left(e^{\bar{x}_e}-\textstyle\sum_{j=0}^{M}\frac{(\bar{x}_e)^j}{j!}\right)\right]=0,
\end{aligned}
$$

where the last equality follows from $\sum_{j=0}^{\infty}\frac{b^j}{j!}=e^b$.

$\square$

Proof of Corollary 1. The proof of Lemma 6 gives that

$$u(K) = u(w^*) - c'(a_e^*)\frac{Q_M}{Q_M^e}.$$

The only piece that varies with M is $\frac{Q_M}{Q_M^e}$. Thus, to show that K becomes increasingly brutal as the threshold increases, it remains only to show that this quotient approaches infinity as M approaches infinity. Recall from the proof of Lemma 6 that $Q_M^e = q_M$; this allows me to rewrite the quotient as follows.

$$\begin{aligned}
\frac{Q_M}{Q_M^e} &= \frac{\sum_{j=0}^{M} q_j}{q_M} \\
&= \frac{e^{-\bar{x}_e}\sum_{j=0}^{M}\frac{(\bar{x}_e)^j}{j!}}{\frac{e^{-\bar{x}_e}(\bar{x}_e)^M}{M!}} \\
&= \left(\frac{M!}{(\bar{x}_e)^M}\right)\left(\sum_{i=0}^{M}\frac{(\bar{x}_e)^i}{i!}\right)
\end{aligned}$$

Now evaluating the limit as $M \longrightarrow \infty$, the first piece in parentheses approaches infinity and the second piece approaches $e^{\bar{x}_e}$. Therefore $\frac{Q_M}{Q_M^e} \longrightarrow \infty$ as $M \longrightarrow \infty$. Then $u(K) \longrightarrow -\infty$ and thus, by assumption (3.27), $K \longrightarrow \underline{s}$ as $M \longrightarrow \infty$. $\square$

Proof of Lemma 7. Let $\theta_j = Pr(y = j|\bar{x}_r, \bar{x}_e)$ and let $\theta_j^e = \frac{\partial}{\partial a_e}\left[Pr(y = j|\bar{x}_r, \bar{x}_e)\right]$. Notice that the agent has no influence over the revenue account; therefore the revenue account has expected value $\bar{x}_r = \hat{x}_r$, and $f_r(\vec{a}) = 0$.

Following the proof strategy of Lemma 6, the contractual payments can be characterized as follows, where $\Theta_i \equiv \sum_{j=-\infty}^{i} \theta_j$ and $\Theta_i^e \equiv \frac{\partial}{\partial a_e} \sum_{j=-\infty}^{i} \theta_j$.

$$u(\tilde{w}) = u(w^*) - c'(a_e^*)\frac{\Theta_M}{\Theta_M^e} \tag{B.55}$$

$$u(K) = u(w^*) + c'(a_e^*)\frac{(1 - \Theta_M)}{\Theta_M^e} \tag{B.56}$$

I want to show that $u(\tilde{w}) \to u(w^*)$ as $M \to -\infty$. As $c'(a_e^*)$ is constant, what I need to show is that $\Theta_M/\Theta_M^e \longrightarrow 0$. Notice that $\theta^e/\theta \longrightarrow -\infty$ implies that $\Theta_M^e/\Theta_M \longrightarrow -\infty$ and thus $1/(\Theta_M^e/\Theta_M) = \Theta_M/\Theta_M^e \longrightarrow 0$. Then all that remains to show is that $\theta^e/\theta \longrightarrow -\infty$; that is, the likelihood ratio of earnings with respect to the cost-cutting action is unbounded below.

Equation (A.14) in Appendix A gives the likelihood ratio LR_y^b for a Skellam-distributed signal $y = x_1 - x_2$ and action a_b. Applying it here, with $x_1 = x_r$, $x_2 = x_e$ and $a_b = a_e$, gives the likelihood ratio below.

$$LR_y^e = \frac{\theta^e}{\theta} = \left[1 - \left(\frac{\bar{x}_r}{\bar{x}_e}\right)^{1/2} \frac{I_{y+1}(2\sqrt{\bar{x}_r\bar{x}_e})}{I_y(2\sqrt{\bar{x}_r\bar{x}_e})}\right]. \tag{B.57}$$

Now I show that $LR_y^e \to -\infty$ as $y \to -\infty$. It is clear that if $I_{y+1}(z)/I_y(z) \to \infty$ as $y \to -\infty$, then $LR_y^e \to -\infty$ as $y \to -\infty$, where $z = 2\sqrt{\bar{x}_r\bar{x}_e}$.

A well known property of the modified Bessel function of the first kind is that $I_{-v}(z) = I_v(z)$ when $v \in \mathbb{Z}$ (Abramowitz and Stegun 1970, p. 375). Because $y \in \mathbb{Z}$, we can write $I_y(z) = I_{-y}(z) = I_n(z)$, where $n = |y|$. Additionally, for $y < 0$, we can write $I_{y+1} = I_{-n+1} = I_{|-n+1|} = I_{n-1}$. Then for $y < 0$, we can rewrite the ratio $I_{y+1}(z)/I_y(z)$ as $I_{n-1}(z)/I_n(z)$, which is equivalent to $I_{y-1}(z)/I_y(z)$ for $y > 0$. Thus, showing that $I_y(z)/I_{y+1}(z) \to \infty$ as $y \to \infty$ is equivalent to showing that $I_{y+1}(z)/I_y(z) \to \infty$ as $y \to -\infty$, which is what I need to prove.

Nåsell (1974) establishes that for all $y \geq -1$ and $z > 0$,

$$1 + \frac{y}{z} < \frac{I_y(z)}{I_{y+1}(z)} \tag{B.58}$$

It's clear that the left-hand side of (B.58) approaches infinity as $y \to \infty$, which implies that the right-hand side does as well. This completes the proof.

$\square$

Proof of Lemma 8. I will prove the lemma by contradiction. Assume that the first-best solution to the revenue-only model, (w^*, a_r^*), can be approximated arbitrarily closely by a contract $(s(x_r), a_r)$, defined as follows.

$$s(x_r) = \begin{cases} \tilde{w} = w^* + \varepsilon & \textit{if } x_r \geq M \\ K & \textit{if } x_r < M \end{cases}$$

$$a_r = a_r^*$$

The agent is punished with penalty payment K for outcomes below threshold M and receives a flat wage, $\tilde{w} > w^*$, for outcomes above M. To make the agent indifferent between the penalty contract and the first-best contract, the penalty contract must give the agent the same expected utility as the first-best contract, and it must be incentive compatible. Thus, $(s(x_r), a_r)$ must satisfy the following two constraints, where $p_i(a_r) \equiv Pr(x_r = i|a_r)$ and $p_i^r(a_r) \equiv \partial/\partial a_r \left[Pr(x_r = i|a_r)\right]$.

$$\sum_{i=0}^{M} u(K)p_i(a_r^*) + \sum_{j=M+1}^{\infty} u(\tilde{w})p_i(a_r^*) = u(w^*) \tag{B.59}$$

$$\sum_{i=0}^{M} u(K)p_i^r(a_r^*) + \sum_{j=M+1}^{\infty} u(\tilde{w})p_i^r(a_r^*) = c'(a_r^*) \tag{B.60}$$

Observe that $p_i^r(a_r) = p_i(a_r) \times LR_r$ by definition of the likelihood ratio $LR_r = \frac{p_i^r(a_r)}{p_i(a_r)}$. Noting that $LR_r = \frac{x_r - \bar{x}_r}{\bar{x}_r} = \frac{x_r}{\bar{x}_r} - 1$, I find that $\sum_{j=0}^{i} p_j^r(a_r)$ can be simplified as follows.

$$
\begin{aligned}
\textstyle\sum_{j=0}^{i} p_j^r(a_r) &= \textstyle\sum_{j=0}^{i} \frac{e^{-\bar{x}_r}(\bar{x}_r)^j}{j!}\left[\frac{j}{\bar{x}_r}-1\right] \\
&= e^{-\bar{x}_r}\left[\frac{(\bar{x}_r)^0}{0!}\left(\frac{0}{\bar{x}_r}-1\right)+\frac{(\bar{x}_r)^1}{1!}\left(\frac{1}{\bar{x}_r}-1\right)+\frac{\bar{x}_r^2}{2!}\left(\frac{2}{\bar{x}_r}-1\right)+\ldots+\frac{(\bar{x}_r)^i}{i!}\left(\frac{i}{\bar{x}_r}-1\right)\right] \\
&= e^{-\bar{x}_r}\Big[(0-1)+(1-\bar{x}_r)+\left(\bar{x}_r-\frac{(\bar{x}_r)^2}{2!}\right)+\left(\frac{(\bar{x}_r)^2}{2!}-\frac{(\bar{x}_r)^3}{3!}\right)+\ldots \\
&\quad \ldots+\left(\frac{(\bar{x}_r)^{i-1}}{(i-1)!}-\frac{(\bar{x}_r)^i}{i!}\right)\Big] \\
&= e^{-\bar{x}_r}\Big[-1+1-\bar{x}_r+\bar{x}_r-\frac{(\bar{x}_r)^2}{2!}+\frac{(\bar{x}_r)^2}{2!}-\frac{(\bar{x}_r)^3}{3!}+\frac{(\bar{x}_r)^3}{3!}-\ldots \\
&\quad \ldots-\frac{(\bar{x}_r)^{i-1}}{(i-1)!}+\frac{(\bar{x}_r)^{i-1}}{(i-1)!}-\frac{(\bar{x}_r)^i}{i!}\Big] \\
&= -\frac{e^{-\bar{x}_r}(\bar{x}_r)^i}{i!} \\
&= -p_i(a_r)
\end{aligned}
$$

Therefore, the term $\sum_{i=0}^{M} p_i^r(a_r^*)$ in equation (B.60) can be replaced with $-p_M$. Notice that because $\sum_{i=0}^{\infty} p_i^r(a_r) = 0$, it must be the case that $\sum_{i=M+1}^{\infty} p_i^r(a_r) = -\sum_{i=0}^{M} p_i^r(a_r) = p_M$.

Let $P_M \equiv \sum_{i=0}^{M} p_i$; then $\sum_{i=M+1}^{\infty} p_i = 1 - P_M$ by definition of a CDF. Noting that $u(K)$ and $u(\tilde{w})$ are fixed values independent of x_r, I rewrite equations (B.59) and (B.60) as follows.

$$u(K)P_M + u(\tilde{w})(1-P_M) = u(w^*) \tag{B.61}$$

$$-u(K)p_M + u(\tilde{w})p_M = c'(a_r^*) \tag{B.62}$$

With some algebraic substitution, equations (B.61) and (B.62) characterize the contractual payments $\tilde{w}$ and K as follows.

$$\tilde{w} = u^{-1}\left(u(w^*) + c'(a_r^*)\frac{P_M}{p_M}\right) \tag{B.63}$$

$$K = u^{-1}\left(u(w^*) - c'(a_r^*)\frac{(1-P_M)}{p_M}\right) \tag{B.64}$$

For the penalty scheme to approximate the first-best solution, it must be that $\tilde{w}$ can be made arbitrarily close to w^*. By equation (B.63), this is equivalent to making $\frac{P_M}{p_M}$ arbitrarily close to zero. But the smallest value attained by $\frac{P_M}{p_M}$ occurs at $M = 0$, where $\frac{P_0}{p_0} = \frac{p_0}{p_0} = 1$. I will show that $\frac{P_M}{p_M} < \frac{P_{M+1}}{p_{M+1}}$ for all $M < E[x_r]$. The proof is by induction. For the base case, it is clear that $\frac{P_0}{p_0} = \frac{p_0}{p_0} < \frac{P_1}{p_1} = \frac{p_0+p_1}{p_1}$ because $p_0 + p_1 > p_1$ and thus $\frac{p_0+p_1}{p_1} > 1 = \frac{p_0}{p_0}$.

Assume the induction hypothesis that for all M less than $E[x^r]$,

$$\frac{P_{M-1}}{p_{M-1}} < \frac{P_M}{p_M}.$$

Now use the recursive property of the Poisson distribution $p_{M-1} = \frac{M}{\bar{x}_r} p_M$ to get

$$\frac{P_{M-1}}{\frac{M}{\bar{x}_r} p_M} < \frac{P_M}{p_M} \iff \frac{P_{M-1}}{p_M} < \frac{P_M}{\frac{\bar{x}_r}{M} p_M},$$

where the equivalence comes from multiplying through by $\frac{\bar{x}_r}{M}$. Now add 1 to both sides to get

$$\frac{P_{M-1} + p_M}{p_M} < \frac{P_M + \frac{\bar{x}_r}{M} p_M}{\frac{\bar{x}_r}{M} p_M}.$$

I use the fact that $\frac{\bar{x}_r}{M} > \frac{\bar{x}_r}{M+1}$ to write

$$\frac{P_{M-1} + p_M}{p_M} < \frac{P_M + \frac{\bar{x}_r}{M} p_M}{\frac{\bar{x}_r}{M} p_M} < \frac{P_M + \frac{\bar{x}_r}{M+1} p_M}{\frac{\bar{x}_r}{M+1} p_M} = \frac{P_M + p_{M+1}}{p_{M+1}}.$$

The above equation shows that $\frac{P_{M-1}+p_M}{p_M} = \frac{P_M}{p_M} < \frac{P_{M+1}}{p_{M+1}} = \frac{P_M + p_{M+1}}{p_{M+1}}$, which completes the proof by iteration.

Having established that $\frac{P_M}{p_M}$ is minimized at $M = 0$ with a value of 1, we revisit equation (B.63) and see that the smallest possible value of $\tilde{w}$ is $\tilde{w} = u^{-1}\left(u(w^*) + c'(a_e^*)\right)$. Then $\tilde{w}$ cannot be made arbitrarily close to w^*, contradicting the opening assertion that the penalty contract could approximate the first-best arbitrarily closely.

□

Proof of Lemma 9. The principal's problem is as follows, letting $p_i(a_r) \equiv Pr(x_r = i|a_r)$ and $p_i^r(a_r) \equiv \partial/\partial a_r\,[Pr(x_r = i|a_r)]$.

$$\begin{array}{rll} \max\limits_{s(x_r),a_r} & \Pi(x_r) - \sum_i s(x_i^r))p_i(a_r) & \text{(OBJ)} \\ \text{subject to} & \sum_i u(s(x_i^r))p_i(a_r) - c(a_r) \geq \bar{U} & \text{(IR)} \\ \text{and} & \sum_i u(s(x_i^r))p_i^r = c'(a_r) & \text{(IC)} \end{array} \tag{B.65}$$

Write program (B.65) in Lagrangian form.

$$\begin{aligned} \mathcal{L}\left(s(x^r), a_r, \lambda_r, \mu_r\right) &= \textstyle\sum_i \left(x_i^r - s(x_i^r)\right) p_i(a_r) \\ &+ \lambda_r[\textstyle\sum_i u(s(x_i^r))p_i(a_r) - c(a_r) - \bar{U}] \\ &+ \mu_r[\textstyle\sum_i u(s(x_i^r))p_i^r(a_r) - c'(a_r)] \end{aligned} \tag{B.66}$$

Differentiating (B.66) with respect to s_i yields equation (3.30) in the lemma. Standard Kuhn-Tucker conditions give that the constant λ_r is positive.

In his Lemma 1, Jewitt (1988) gives a clever proof that $\mu_r > 0$. I repeat his argument here for completeness. Any (s, a_r) which solves program (B.65) must satisfy equations (3.30) and the (IC) constraint. Equation (3.30) can be rearranged as

$$p_i^r(a_r) = \left[\frac{1}{u'(s_i)} - \lambda_r\right] \frac{p_i(a_r)}{\mu_r}.$$

Plugging this expression into equation the (IC) constraint and rearranging gives

$$\sum_i u(s_i) \left[\frac{1}{u'(s_i)} - \lambda_r\right] p_i(a_r) = c'(a_r)\mu_r. \tag{B.67}$$

Now I take the expectation of both sides of (3.30) and simplify, employing the fact that

$\sum_i p_i^r = 0$.

$$
\begin{aligned}
& \textstyle\sum_i \frac{1}{u'(s_i)} p_i(a_r) = \sum_i \left[\lambda_r + \mu_r \frac{p_i^r(a_r)}{p_i(a_r)}\right] p_i(a_r) \\
\iff \quad & \textstyle\sum_i \frac{1}{u'(s_i)} p_i(a_r) = \lambda_r + \mu_r \sum_i p_i^r(a_r) \\
\iff \quad & \textstyle\sum_i \frac{1}{u'(s_i)} p_i(a_r) = \lambda_r \\
\iff \quad & E\left[\frac{1}{u'(s_i)}\right] = \lambda_r
\end{aligned}
\tag{B.68}
$$

Recall that $E\Big[\big(g(y) - E\left[g(y)\right]\big)\big(h(y) - E\left[h(y)\right]\big)\Big]$ is the covariance between $g(y)$ and $h(y)$. Because the utility function $u(\cdot)$ is an ordinal representation and unaffected by affine transformations, we can normalize $u(\cdot)$ such that $E[u(s_i)] = 0$. Then equation (B.67) can be written

$$
\begin{aligned}
& \sum_i \left[u(s_i) - 0\right]\left[\frac{1}{u'(s_i)} - \lambda_r\right] p_i(a_r) = c'(a_r)\mu_r \\
\iff & E\left[\Bigg(u(s_i) - E\left[u(s_i)\right]\Bigg)\left(\frac{1}{u'(s_i)} - E\left[\frac{1}{u'(s_i)}\right]\right)\right] = c'(a_r)\mu_r,
\end{aligned}
\tag{B.69}
$$

where the left-hand side is the covariance between $u(s_i)$ and $1/u'(s_i)$. Signing this covariance will help us sign μ_r. $u(\cdot)$ is increasing in its argument. The agent's risk aversion gives $u''(\cdot) < 0$, so $u'(\cdot)$ is decreasing in its argument and thus $1/u'(\cdot)$ is increasing in its argument. So $u(s_i)$ and $1/u'(s_i)$ are both increasing in s_i and thus $cov\left(u(s_i), 1/u'(s_i)\right) > 0$, so the left hand side of (B.69) is positive. The agent's effort aversion gives $c'(a_r) > 0$, so for (B.69) to hold, μ_r must be positive.

Armed with $\mu_r > 0$, return to examining equation (3.30). $\frac{x_i^r - \bar{x}_r}{\bar{x}_r}$ is increasing in i, so with $\lambda_r, \mu_r > 0$, the right-hand side of (3.30) is increasing in i. This implies that the left-hand side is also increasing in i. We know already that $1/u'(s_i)$ must be increasing in its argument, so for the left-hand side to be increasing in i, s_i must be increasing in i, which is what the lemma claims.

$\square$

BIBLIOGRAPHY

Abramowitz, Milton, and Irene A Stegun. 1970. *Handbook of mathematical functions with formulas, graphs, and mathematical tables.* Vol. 55. Applied Mathematical Series. New York: Dover.

Alzaid, Abdulhamid A, and Maha A Omair. 2010. On the Poisson difference distribution inference and applications. *Bulletin of the Malaysian Mathematical Sciences Society* 8 (33): 17–45.

Amershi, Amin H, Rajiv D Banker, and Srikant M Datar. 1990. Economic sufficiency and statistical sufficiency in the aggregation of accounting signals. *Accounting Review* 65 (1): 113–130.

Carter, Mary Ellen, Edie Hotchkiss, and Mahdi Mohseni. 2018. Payday before mayday: CEO compensation contracting for distressed firms. Working Paper.

Christensen, P, and G Feltham. 2003. *Economics of Accounting.* Vol. I. Information in markets. Boston: Kluwer Academic Publishers (also New York: Springer).

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

Demski, Joel S, John C Fellingham, Yuji Ijiri, and Shyam Sunder. 2002. Some thoughts on the intellectual foundations of accounting. *Accounting Horizons* 16 (2): 157–168.

Drake, Katharine D, Ellen Engel, and Melissa Martin. 2018. Money for nothing? Using expectations of loss persistence to examine CEO cash compensation in loss firms. Working Paper.

Eckbo, B Espen, Karin S Thorburn, and Wei Wang. 2016. How costly is corporate bankruptcy for the CEO? *Journal of Financial Economics* 121 (1): 210–229.

Gaver, Jennifer J, and Kenneth M Gaver. 1998. The relation between nonrecurring accounting transactions and CEO cash compensation. *The Accounting Review* 73 (2): 235–253.

Gigler, Frank B, and Thomas Hemmer. 2001. Conservatism, optimal disclosure policy, and the timeliness of financial reports. *The Accounting Review* 76 (4): 471–493.

Gilson, Stuart C, and Michael R Vetsuypens. 1993. CEO compensation in financially distressed firms: an empirical analysis. *The Journal of Finance* 48 (2): 425–458.

Graham, John R, Campbell R Harvey, and Shiva Rajgopal. 2005. The economic implications of corporate financial reporting. *Journal of Accounting and Economics* 40 (1): 3–73.

Hemmer, Thomas, Oliver Kim, and Robert E Verrecchia. 2000. Introducing convexity into optimal compensation contracts. *Journal of Accounting and Economics* 28 (3): 307–327.

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

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

Hughes, John S, Li Zhang, and Jai-Zheng James Xie. 2005. Production externalities, congruity of aggregate signals, and optimal task assignments. *Contemporary Accounting Research* 22 (2): 393–408.

Jewitt, Ian. 1988. Justifying the first-order approach to principal-agent problems. *Econometrica* 56 (5): 1177–1190.

Kim, Son Ku, and Yoon S Suh. 1991. Ranking of accounting information systems for management control. *Journal of Accounting Research* 29 (2): 386–396.

Leone, Andrew J, Joanna Shuang Wu, and Jerold L Zimmerman. 2006. Asymmetric sensitivity of CEO cash compensation to stock returns. *Journal of Accounting and Economics* 42 (1-2): 167–192.

Matějka, Michal, Kenneth A Merchant, and Wim Van der Stede. 2009. Employment horizon and the choice of performance measures: empirical evidence from annual bonus plans of loss-making entities. *Management Science* 55 (6): 890–905.

Mirrlees, James. 1979. The implications of moral hazard for optimal insurance. Working Paper, Nuffield College, Oxford.

———. 1999. The theory of moral hazard and unobservable behaviour. Part I. *The Review of Economic Studies* 66 (1): 3–21.

Nåsell, Ingemar. 1974. Inequalities for modified Bessel functions. *Mathematics of Computation* 28 (125): 253–256.

Rogerson, William P. 1985. The first-order approach to principal-agent problems. *Econometrica* 53 (6): 1357–1367.

Sannikov, Yuliy. 2008. A continuous-time version of the principal-agent problem. *The Review of Economic Studies* 75 (3): 957–984.

Shaw, Kenneth W, and May H Zhang. 2010. Is CEO cash compensation punished for poor firm performance? *The Accounting Review* 85 (3): 1065–1093.

Skellam, John G. 1946. The frequency distribution of the difference between two Poisson variates belonging to different populations. *Journal of the Royal Statistical Society. Series A (General)* 109 (Pt 3): 296.

Income Statement

Revenues

$x_{r1} \sim Poisson(\hat{x}_{r1} + f_{r1}(\vec{a}))$

$x_{r2} \sim Poisson(\hat{x}_{r2} + f_{r2}(\vec{a}))$

$\vdots$

$x_{rm} \sim Poisson(\hat{x}_{rm} + f_{rm}(\vec{a}))$

$\vec{a}$

Total revenues $= x_r \sim Poisson\left(\sum_{i=1}^{m}(\hat{x}_{ri} + f_{ri}(\vec{a}))\right)$

Expenses

$x_{e1} \sim Poisson(\hat{x}_{e1} + f_{e1}(\vec{a}))$

$x_{e2} \sim Poisson(\hat{x}_{e2} + f_{e2}(\vec{a}))$

$\vdots$

$x_{en} \sim Poisson(\hat{x}_{en} + f_{en}(\vec{a}))$

Total expenses $= x_e \sim Poisson\left(\sum_{j=1}^{n}(\hat{x}_{ej} + f_{ej}(\vec{a}))\right)$

Earnings $= y \sim Skellam\left(\sum_{i=1}^{m}(\hat{x}_{ri} + f_{ri}(\vec{a})); \sum_{j=1}^{n}(\hat{x}_{ej} + f_{ej}(\vec{a}))\right)$

Figure 1: Income statement with Poisson-distributed accounts. A vector of managerial actions $\vec{a}$ influences the firm's m revenue accounts, $x_{r1}, ..., x_{rm}$, and n expense accounts, $x_{e1}, ..., x_{en}$. Each account follows a Poisson distribution parameterized by its expected value, $\hat{x} + f(\vec{a})$, where $\hat{x}$ is exogenous and $f(\vec{a})$ is the account's change in expected value from managerial actions (where $f(\vec{a})$ may equal zero). Total revenues and total expenses also follow Poisson distributions with parameters simply equal to the sum of the revenue and expense account parameters, respectively. Earnings are equal to total revenues minus total expenses and follow a Skellam distribution whose first and second parameters are the parameters from total revenues and total expenses.

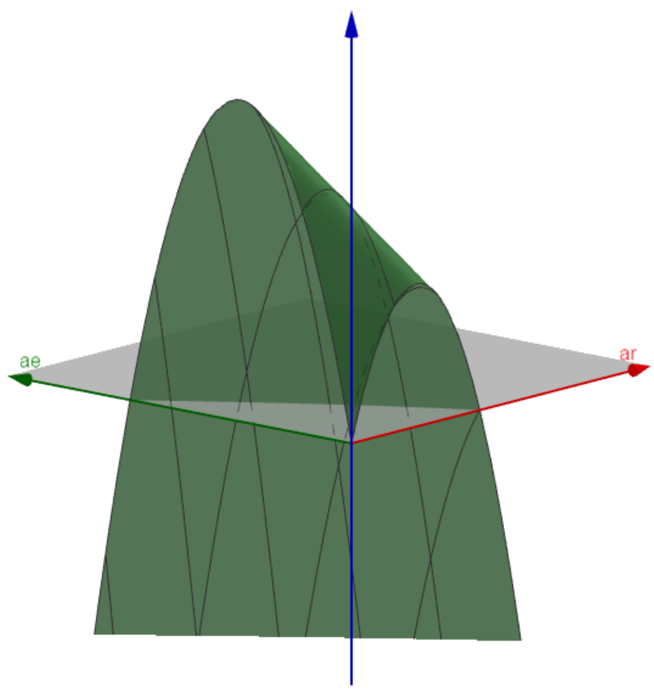

Figure 2: Corner solution with $\boldsymbol{\pi_b = 0}$. The principal's expected residual payoff $(\Pi - E(s))$ is graphed over actions a_e and a_r. Because a_e is cheaper to implement than a_r, a linear profit function and linear cost of effort function result in a corner solution in which the principal sets $a_r = 0$ and maximizes only over a_e.