Aggregation and Convexity in the Provision of Dynamic Incentives

Thomas Hemmer*

November 9, 2022

Abstract

In this paper I identify an alternative preference structure that preserves most of the cherished simplicity of the formulation of the Principal-Agent problem pioneered by Holmström and Milgrom (1987). The main advantage of my approach is in relation to the structure of the optimal contract: it adds a convex component to their optimal linear contract. This provides new opportunities to revisit empirical predictions and studies based off of their linear formulation and to demonstrate how the empirical irregularities may be at least partially explained by this one additional component identified here.

*Preliminary, incomplete and rough draft. I am grateful to participants in UIC and Purdue's Accounting Research Conferences and in the Accounting Workshops at University of British Columbia, Chapman University Columbia University and UCSD for comments on earlier drafts Please do not circulate further or quote.

1 Introduction

The seminal paper by Holmström and Milgrom (1987) established elegantly that more complex contracting problems may well have the simpler solutions. First, by allowing the agent full control over the output distribution as opposed to limit his control to one or a few central parameters, they showed that when it comes to implementing a particular distribution, the principal's hands are tied: the contract that implements a given distribution is unique and thus not subject to the complications introduced by the standard minimization needed to find the optimal contract that implements an action in more limited choice situations, such as Grossman and Hart (1983). While the contract-action combination that solves the principal's over-arching problem of expected utility maximization of course may still not have a unique solution, finding a contract-action pair that does solve the principal's problem does arguably get significantly simpler when all contract-action pairs are unique.

Second, Holmström and Milgrom (1987) proceeded to demonstrate that when the principal and agent have CARA preferences over aggregate consumption net of the aggregate cost of effort as denominated in the units of compensation, a solution to a dynamic multi-period extension of this problem to one where the agent observes past performance prior to his next action is for the principal to implement the same action using the same unique contract each and every period. Moreover, implementing this solution becomes particularly simple: the contract can be written as a linear combination of the aggregate balances of a set of enumeration accounts that track the (relative) frequency of a particular outcome being realized. In the limiting case of continuos time, these balances become normal distributed yielding the added bonus of an easy to obtain closed form approximate solution to the optimal dynamic contract written on such aggregate performance measures.

The key to the latter part is of course stationarity and this is where the restrictions on the parties preferences are crucial. As Holmström and Milgrom (1987) show, their particular format lends itself to decomposing the principal's multi-period problem into as many one-period problems as there are sub-periods. Moreover, the solution to each of these problems is independent of the solution to any of the remaining problems and thus is the same as the solution to the single period problem. Without this feature, the solution becomes path dependent and the aggregation (and linearity) result no longer holds. Those that have exploited the tractability of this framework, weather in its fundamental form or relying on the so-labeled "ad-hoc principle," have thus also been bound by the particular multiplicatively-separable, negative-exponential preference set-up.

While the linearity in normal distributed performance measures is a boon to tractability that has been instrumental in extending P-A models in ways that otherwise are off limits, as always there are no (entirely) free lunches. Some argue that while, when it comes to the nature of empirically observed contracts, linearity may be a good first (local) approximation, it is not the whole story. Many contracts have significant convex features such as options. Also, the unlimited downside not just in utility space but in terms of the actual cash transfers that the linear contracts entertain in a normally distributed world does not seem to be entirely descriptive either. Wealth concerns and legal limitations may be somewhat incompatible with this from a more descriptive vantage point.

Perhaps unsurprisingly, then, there have been ongoing efforts to develop an alternative framework that, while still offering the coveted tractability, also could provide for the analysis of settings where linearity itself limits the scope of the issues that can be satisfactorily addressed. Edmans and Gabaix (2011), for example, modified the assumptions of Holmström and Milgrom (1987) in several significant ways to this end. First, they allow for any preference representation consistent with A1 in Grossman and Hart (1983).[1] Second, they restrict the underlying uncertainty to a particular class of probability distributions over which the agent has only limited control. Third, the order of play is partially reversed - in their setting the agent observes the state of nature *before* choosing his action. Fourth, the consequences of the agent's action(s) is assumed to be deterministic such that period output is simply the sum of resolved uncertainty and the agents (informed) choice.

[1] The preference representation of Holmström and Milgrom (1987) is clearly a special case hereof.

The first deviation clearly improves generality. The other three are designed to ensure that the control problem remains rich enough that the uniqueness feature of Holmström and Milgrom (1987) is not lost while gaining some control over the nature of equilibrium output distributions. What *is* lost by Edmans and Gabaix (2011), however, is the all important stationarity result. Because of this and because little can be said in their setting without this critical property, they eventually resort to imposing a set of assumptions that guarantees that the most productive action is *always* strictly preferred and thus will always be implemented by the principal.

The proposed benefits of all of this is about the nature of the tractable contract that implements this "most productive action." Such contracts tend to be convex and thus lend themselves to types of inquiry contract linearity may preclude. The drawbacks are not insignificant, however. The reverse order of play is not merely a technicality but a very real and potentially serious limitation of this analysis. If it is to be thought of as economically relevant, it must be that it is a fair characterization of a meaningful sub-set of firms in the economy. However, if firms generally can hire a manager that will be able to observe their resolution of uncertainty perfectly at any given point in time, there must be an significant role and market for monitors beyond what is present in standard P-A set-ups.

In addition, the strong assumptions, while sufficient to guarantee the "stationarity" in their setting, deviations two to four above, are really not necessary. Holmström and Milgrom (1987) extends easily to alternative preference representations and, in particular, to convex tractable contracts, when it can be assumed that the same action is dominant in each period regardless of history. This is true whether a particular action is assumed dominant directly or indirectly such as via, for example, the "High Effort Principle" proposed by Edmans and Gabaix (2011). As of now, however, no framework exists that avoids either the multiplicatively separable negative exponential (CARA) specification and the resulting linearity of Holmström and Milgrom (1987) or, in lieu of this, the drawbacks of additional assumptions and restrictions then needed to make the contract tractable. This paper establishes one such arguably straight-forward extension of Holmström and Milgrom (1987) that achieves both.

The observation behind the extension developed here is a relative simple one but one that appears to have been entirely ignored in the literature: CARA somewhat loosely implies that the pecuniary loss (the "risk-premium" associated with a local "fair" permutation in *cash*-space is insensitive to the level of wealth. But, as made precise later, there similarly exists a particular CRRA utility function for which the pecuniary loss of a particular fair permutation in *utility*-space is also independent of wealth. Accordingly, as I will show generally, just as the CARA set-up of Holmström and Milgrom (1987) where cost of effort is denominated in cash leads to optimal effort-stationarity over time, so does this particular CRRA set-up when effort costs are denominated in utiles. The latter is, of course, consistent with the standard additively separable utility function underlying classical papers such as Holmström (1979) a.m.o.

The benefits here are that optimal contracts, while remaining highly "tractable" are strictly convex. Indeed, the model set-up presented and analyzed here extends the original Holmström and Milgrom (1987) framework to such cases without giving purchase on the simplicity of their solution. In particular, convexity and simplicity is achieved without any of the baggage of Edmans and Gabaix (2011). Thus, because it is possible to obtain simple closed form solutions to the contracting problem that invites convexity of optimal aggregate performance-based contracts, the relevant comparative statics that can be obtained here can also be compared and contrasted directly with those obtained from Holmström and Milgrom (1987).

I proceed as follows. First I introduce the model modifications that will preserve all features of Holmström and Milgrom (1987) except the linearity. Second I will introduce the specific structure and establish formally the resulting time- and outcome-independence of the optimal sub-period effort-contract pairing when the optimal contract is represented in utility space. I will then elaborate on the properties of the optimal contract when evaluated in cash compensation space. Finally, I conclude by offering some suggestions about potentially interesting applications of this framework.

2 A Single-Period Model

Key to the aggregation result of Holmström and Milgrom (1987) was to first establish for their multiplicatively separable CARA preference representation, that the solution to the principal's problem is independent of the agent's personal wealth or, equivalently, of the RHS of the standard IR-constraint (Theorem 2). This is, of course, critical to the simplicity of the multi-period extension because net wealth accumulated during past periods (or expected to be accumulated in the future) thus has no impact on the present. The principal's problem for each sub-period therefore can be solved individually as one-period problems and all have the exact same solution.

The approach I take here is similar in terms of sequencing, but the nature of the additively separable preference representation I employ leads to a fundamental difference between the single and multi-period models that is instructive. To that end, assume that the principal is risk-neutral and cares only about terminal aggregate output, Π, net of the agent's compensation. The final output depends on the agent's actions during the contracting horizon but is not directly observable/contractible during the relevant time-frame. Instead, the principal observes a set of $N+1$ informative signals, $x_j \in x$, $j = 0, ..., N$. The signals here are ordered from lowest to highest and wlog I normalize $x_0 = 0$. The agent's action choice is of full dimension in the sense that he chooses directly the probability of x_i, $p_i \in p$, $i = 1, ..., N$, with the constraint that $p_0 = 1 - \sum_{i=1}^{N} p_i > 0$.

As mentioned previously, the agent here is risk and effort averse as represented by the following additively separable utility function defined over end-of-horizon consumption, y,

$$H(y, p) = u(y) - v(p),$$

where $u' > 0$ and risk aversion implies $u" < 0$, and the personal cost associated with implementing "action" p. Using $v_i(p)$ and $v_{ik}(p)$ to denote the partial derivative of $v(p)$ w.r.t. p_i and the cross-partial of $v_i(p)$ w.r.t. p_k respectively, I assume that $v_i(p) > 0$ and $v_{ii}(p) > 0$, $i = 1, ..., N$ while for simplicity I assume that $v_0(p) = 0$ $\forall p$.

With this, it is straight forward to confirm that Theorem 3 of Holmström and Milgrom (1987) the uniqueness of the contract that implements a particular p extends readily to the additively separable case considered here. The agent's expected utility under a given contract, $s(x_i)$, is given a

$$\theta \equiv \sum_{i=0}^{N} u(s(x_i))\, p_i - v(p) \tag{1}$$

and assuming p is the interior,[2] we have the agent's first-order condition as

$$u(s(x_i)) - u(s(x_0)) = v_i(p) - v_0(p) \tag{2}$$

where uniqueness then follows from strict convexity of $v(p)$. Then using (1) and (2) we have

$$u(s(x_i)) = \theta + v(p) + (1 - p_i)\, v_i(p) - \sum_{k \neq i} p_k v_k(p).$$

Letting $z(.)$ be the inverse of $u(y)$ such that $z(u(y)) \equiv y$, the standard and familiar formulation of the principal's (first-order) one-period problem thus becomes

$$\max_{p} \quad p'x - \sum_{i=0}^{N} p_i z\left(\theta + v(p) + (1 - p_i)\, v_i(p) - \sum_{k \neq i} p_k v_k(p)\right)$$

with first-order conditions

$$x_i = (s_i - s_0) + \sum_{i=0}^{N} p_i G'\left(\theta + v(p) + (1 - p_i)\, v_i(p) - \sum_{k \neq i} p_k v_k(p)\right) \times \left[(1 - p_i)\, v_{ii} - \sum_{k \neq i} p_k v_{ki}(p)\right].$$

Obviously, then, here the role of the expected utility, θ, depends on the cross-partials of the cost function. However, for the arguably neutral case where $v_{ki}(p) = 0$, $\forall k \neq i$, all terms on

[2] I will introduce simple assumptions below that guarantee this.

the RHS are increasing in θ due to the convexity of $G(\cdot)$. Accordingly, the marginal cost of p_i is then increasing in θ for all p_i and the optimal p_i, $i = 1, ..., N$, is therefore is decreasing in θ for *any* concave $u(\cdot)$.

The broader message is, however, that there is no (obvious) counterpart to Theorem 4 in Holmström and Milgrom (1987) in the case of additively separable preferences such as those introduced above: the optimal action generally depends directly on the agent's wealth and his expected utility requirement. While that may at first seem to rule out getting tractability in the multi-period extension outside of out-right exogenously imposing that one action is always the dominant one, it is worth noting that the assumptions in Holmström and Milgrom (1987) are sufficient to guarantee that the optimal contract is a simple solvable function of aggregate performance. They are not necessary, however, as I will proceed to establish.

Before doing so, the following insight is helpful. Let w will be some given level of wealth of the agent and consider now offering the agent a (cash) lottery, $\widetilde{\epsilon}$, that is actuarially fair in *utility space.* That is

$$E\left[u\left(w+\widetilde{\epsilon}\right)\right] = u\left(w\right).$$

Let then $\widetilde{\xi}$ represent the induced variation in the agent's utility by this lottery so that for any given realization of $\widetilde{\epsilon}$, the corresponding realization of ξ is determined as

$$\xi \equiv u\left(w+\epsilon\right) - u\left(w\right).$$

Now, since the agent is risk-averse, the exist a strictly positive number, δ, such that

$$E\left[z\left(u\left(w\right)+\widetilde{\xi}\right)\right] = w+\delta,$$

where δ then is the actuarial cash value of the expected-utility-neutral lottery $\widetilde{\epsilon}$. Using a

Taylor-expansion of the LHS, we have

$$E\left[z\left(u\left(w\right)\right)+\widetilde{\xi}z'\left(u\left(w\right)\right)+\frac{1}{2}\widetilde{\xi}^{2}z"\left(u\left(w\right)\right)\right]=w+\delta$$

or

$$\frac{1}{2}var\left(\widetilde{\xi}\right)z"\left(u\left(w\right)\right)=\delta. \tag{3}$$

This leads directly to the following Lemma:

Lemma 1 *The cash value of a lottery acceptable to a risk-averse agent is independent of the agent's current wealth iff that agent has a power utility function where the power is one-half so that*

$$u\left(w\right)=\left(\frac{w-a}{b}\right)^{1/2}, \qquad a<w,\ b>0.$$

Proof. It follows from (3) that the actuarial value for a given $var\left(\widetilde{\xi}\right)$ depends only on $z"\left(\cdot\right)$. Accordingly, iff $z"\left(u\left(w\right)\right)=k>0$, where k is some constant and the last inequality follows from the agent being strictly risk-averse, δ is independent of the agent's level of (expected) utility. Then, integrating twice over both sides of $z"\left(u\left(w\right)\right)=k$ recovers

$$z\left(u\left(w\right)\right)=a+b\left(u\left(w\right)\right)^{2}, \quad b>0,$$

and the lemma then follows. ■

As I will proceed to show, this feature is a key component for the optimal action to be time- and outcome invariant in the case of additively separable preferences. For the remainder of this paper I therefore rely on the following specific preference representation for the agent:

$$H\left(y,p\right)=\left(\sum_{t=1}^{T}y_{t}\right)^{1/2}-\sum_{t=1}^{T}v\left(p_{t}\right), \quad T>1. \tag{4}$$

3 Multi Sub-Period Extension Considerations

To extend the single-period model to a continuos time/outcome version, I proceed in two steps: I first identify the fundamental problem involved in multi sub-period extensions of the basic model when deviating from the multiplicatively separable CARA preference structure of Holmström and Milgrom (1987). I then identify a production and measurement structure set-up that will suffice in terms of eliminating the complications at hand while, at the same time, not change the nature of the principal's dynamic problem and also enrich the nature of the contract that solves the principal's problem.

The first point to be made here is that, unlike in the multiplicatively-separable case, the fact that the cost of implementing a particular action is (also) wealth independent in the additively separable "square root" case pursued here does *not* logically imply that effort is also necessarily time and outcome independent.

Lemma 2 *For the multi-period version of the model introduced here, it is not the optimal strategy to make the agent's future actions independent of past outcome realizations.*

Proof. The result follows directly from Proposition 2 in Hemmer (2017a). To sketch a proof here consider the simple case with just two sub periods and binary performance. Suppose period two effort is independent of period 1 output. Based on Lemma 1 it is straight forward to verify that $u\left(\overline{s}_1\left(x_{11}\right)+\overline{s}_2\left(x_{2i}\right)\right)-u\left(\overline{s}_1\left(x_{10}\right)+\overline{s}_2\left(x_{2i}\right)\right)=u\left(\overline{s}_1\left(x_{1i}\right)+\overline{s}_2\left(x_{21}\right)\right)-u\left(\overline{s}_1\left(x_{1i}\right)+\overline{s}_2\left(x_{20}\right)\right),\ i=0,1$. Now consider modifying this contract so that $\widetilde{s}_2\left(x_{10},x_{20}\right)\equiv\overline{s}_2\left(x_{20}\right)-\delta s_0$, $\widetilde{s}_2\left(x_{10},x_{21}\right)\equiv\overline{s}_2\left(x_{21}\right)+\delta s_1$, $\widetilde{s}_2\left(x_{11},x_{20}\right)\equiv\overline{s}_2\left(x_{20}\right)+\delta s_0$ and $\widetilde{s}_2\left(x_{11},x_{21}\right)\equiv\overline{s}_2\left(x_{21}\right)+\delta s_1$, where s_0 and s_1 are positive constants and $\delta\in R$. First verify that s_0 and s_1 can be constructed such that both the derivative of the expected compensation and the derivative of expected second period effort w.r.t. δ are zero at $\delta=0$. Second, then verify that when s_0 and s_1 are such constructed, the derivative of first period effort w.r.t. δ evaluated at $\delta=0$ is strictly positive. By adding some outcome based variation in the restricted contract $\overline{s}_t\left(x_t\right)$, the principal can achieve the same first period incentives at a lower cost as the added incentives provided by second period work-load variations allow the principal to reduce the compensation risk associated with first period output. ■

While this difference between the multiplicative and additive separable representations is perhaps interesting in its own right, the desire to make future actions conditioned on past outcomes is of course detrimental to the ability to write the optimal contract as a simple function of aggregate performance. At first glance it would seem that adding structure to prevent this natural economic demand for path dependent pay and economic activity from playing out would be somewhat unappealing and not that much different from assuming that there is a dominant action to be implemented each and every sub-period. A more reasonable way of looking at the difference between this and the result of Holmström and Milgrom (1987) is to note that production volatility generally is considered as a negative for a variety of practical reasons.

One clear such reason is that it requires costly slack capacity being kept in reserve. While the cost of capacity is not relevant in their formulation and thus can be safely ignored, it is the implicit assumption that the (marginal) cost of capacity is zero (or sufficiently low). That implicit assumption is one reason the optimal solution is path-dependent here. While adding a "sufficiently large" capacity costs could recover "effort stationarity," I'll pursue a much simple approach that achieves the same objective here. Rather than acquiring capacity, I'll simply assume that the principal at the start of the contracting horizon must acquire a (for the horizon) fixed technology that must be compatible, as to be defined in the following, with the action(s) chosen by the agent.

As in Holmström and Milgrom (1987), first I make a distinction between value creation and measures thereof. The publicly observable construct x introduced above is thus, hereafter, simply a measure that reflect value creating activities like an accounting measure. The actual value created in each sub-period, π_t, is not assumed to be observable during the relevant horizon. It's expected value, Π_t, however, depends on both the firms chosen technology and the agent's actions. Specifically, let the pre-determined technology for period

t be represented by a $N+1$ element vector ϕ_t with $\phi_{tj} \subseteq [0,1]$. Further assume that for $p_{ti} = \phi_{ti}$, $\Pi_t(p_t, \phi_t) = \sum_{i=1}^{N} \pi_{ti} p_{ti}$ while zero otherwise. In other words, if the principal wishes to increase p_{ti}, $i = 1, ...N$, for each p_{ti} he must also install a higher ϕ_{ti} at the *start* of the contracting horizon. With this added structure I can state the following Lemma:

Lemma 3 *Let the agent's utility function be given by (4). Then the optimal solution is for the principal is to implement the same* $\{p_t, \phi_t\}$ *pair* $\forall t = 1, ..., T$.

Proof. *When the action cannot be contingent on realized performance, absence of wealth-effects is sufficient for the optimal action to be time independent as well.* ■

At first glance the assumption on the link of the technology, productive action and output may seems somewhat heavy-handed. There are several counter-points to that, however. First note that this assumption is not sufficient to ensure time-independence of the optimal effort alone - this also takes the properties of the utility function (4). In Holmström and Milgrom (1987), their assumption that the agent's (convex) cost of effort is denominated in cash rules out any potential benefits of using future actions to incentivize current actions. Adding then CARA ensures that the principal implements the same particular action in every sub-period. In the additively separable setting I analyze here, as long as *some* feature of the model, such as a "sufficiently" convex capacity cost or, much simpler, just having to pre-commit to the type of technology I rely on, renders outcome dependent actions unattractive, time independence is assured by(4).

Second, from an applied perspective, and the exercise here is aimed at producing applicability, nothing is lost relative to Holmstrom and Milgrom (1987). Both the optimal action and, thus, the properties of the optimal contract, of course, still depend on the particular exogenous properties of the problem and as such are amenable to the full set of comparative statics. That contrasts sharply with the approach of Edmans and Gabaix (2011) that requires a singular preferred action be present regardless of the preferences of the agent(!) and the parameters of the environment in which the agent is operating. From these vantage points, the assumption that the actions match the technology/strategy chosen by the principal, seems quite benign.[3]

4 Brownian Approximation

With the additional structure introduced in the previous section the multi-period contract has the same qualitative property as that obtained by Holmström and Milgrom (1987): it can be written as a simple function of aggregate account balances at the end of the contracting horizon. The only difference is that in the case analyzed here the contract is linear in the account balances in *utility space* rather than in terms of actual cash disbursements. As a result, here the contract is actually convex in cash space due to the convexity of the inverse utility function. Importantly, the particular utility specification behind the results obtained lends itself well to obtaining simple-to-calculate closed form convex contracts using the same "Brownian Approximation" pursued in the latter part of Holmström and Milgrom (1987). For simplicity I focus on approximating the simple binomial case, as extensions are straightforward and left to the reader.

To extent the single-period model to a continuos time/outcome version, I also rely heavily on the methodology provided by Hellwig and Schmidt (2002) with only minor departures to ensure consistency with the alternative preference representation and added structure I have introduced above. I proceed in two steps: I first subdivide the one-period model into a discrete time version with m sub-periods of length $\Delta = 1/m$ to identify specific conditions under which the optimal actions for all sub-period are the same. Subsequently I proceed to show that, in the limit as $\Delta \to 0$, the solution becomes indistinguishable from that of a one-shot model where the principal incentivizes the agent to take a costly action once that determines the mean and variance of an (approximately) normally distributed performance measure via a contract that is a linear function of this performance measure in utility space

[3] Note again that here the assumptions are sufficient to deliver the desired tractability. In both Holmstrom and Milgrom (1987) and Edmans and Gabaix (2011), the assumptions are necessary.

and thus convex in cash-space.

By restricting attention to the case of $N = 1$ allows me to simplify the probability structure by letting $p_1 \equiv p$ and $p_0 \equiv 1 - p$. The fundamental uncertainty embedded in output, then, is characterized by the exogenous probability $\widehat{p}$ (> 0) which is the probability of a positive outcome when no (unobservable) effort is provided. For the base-case where $m = 1$ let $\kappa \equiv x_1 - x_0$, where $x_1 > \widehat{\mu} > x_0$, and $\widehat{\mu}$ is the the expected output associated with $\widehat{p}$. Denote by $\widehat{\mu}^{\Delta} \equiv \Delta\widehat{\mu}$, $x_1^{\Delta} \equiv \Delta^{1/2}x_1$ and $x_0^{\Delta} \equiv \Delta^{1/2}x_0$, so that $\kappa^{\Delta} = \Delta^{1/2}\kappa$, the values of the corresponding variables in a given sub-period of length Δ. Let the expected output associated with $\widehat{p} \equiv (\widehat{\mu} - x_0)/\kappa$. From the prior sections it already follows that as in Holmström and Milgrom (1987) an optimal solution to the principal's problem is to induce the agent to select the same action, denoted p^{Δ} $(\geq \widehat{p})$, in each of the m sub-periods.

With only a slight departure from Hellwig and Schmidt (2002), the personal cost of p^{Δ} to the agent is here assumed to take the form

$$c^{\Delta}\left(p^{\Delta}\right) = \Delta c\left(\frac{p^{\Delta} - \widehat{p}}{\Delta^{1/2}}\right).$$ [4]

This ensures that the cost of enhancing expected performance *beyond* $\widehat{\mu}$ is independent of the number of sub-periods, m, which of course also implies that the cost of implementing a particular μ is independent of m here. Moreover, suppose the principal wants to implement a particular $\mu \geq \widehat{\mu}$ for the entire horizon by implementing $\mu^{\Delta} \equiv \Delta\mu$ in each of the m sub-periods, he must implement in each sub-period $p^{\Delta} = \widehat{p} + \Delta^{1/2}(p - \widehat{p})$. The cost of that per sub-period is simply, then, $\Delta c(p - \widehat{p})$ and the total cost for the entire horizon thus is simply equal to $c(p - \widehat{p})$ independent of m.[5] Further, for the principal to implement a particular

[4] Hellwig and Schmidt (2002) rely on the cost-function

$$c^{\Delta}\left(p^{\Delta}\right) = \Delta c\left(\widehat{p} + \frac{p^{\Delta} - \widehat{p}}{\Delta^{1/2}}\right).$$

[5] Please note that the notation $\mu^{\Delta} \equiv \Delta\mu$ is a deviation from the similar notation in Hellwig and Schmidt (2002). In their notation, μ^{Δ} represents the same construct as μ does in this paper.

p^Δ or, equivalently, a particular μ^Δ in any given sub-period, the contract must be such that the desired p^Δ solves

$$u\left(x_1^\Delta\right) - \left(x_0^\Delta\right) = \Delta^{1/2} c' \left(\frac{p^\Delta - \widehat{p}}{\Delta^{1/2}}\right).$$

Since the marginal benefit in each sub-period is $\kappa^\Delta = \Delta^{1/2}\kappa$ here, the (first-best) marginal cost/benefit trade-off is also independent of the number of sub-periods with this particular structure.

Consider now the utility-performance-sensitivity in any given sub period,

$$\beta^\Delta \equiv \left(u\left(x_1^\Delta\right) - \left(x_0^\Delta\right)\right)/\Delta^{1/2}\kappa = c'/\kappa. \tag{5}$$

Because β^Δ thus is also independent of m I'll drop the superscript and simply refer to this construct as β. Now, let with a bit abuse of notation convention, $\mu^\nabla \equiv \mu - \widehat{\mu}$. The second-best cost of implementing a particular μ over and above the first-best cost of procuring μ directly while providing the agent with an expected utility of θ then can be found as

$$\begin{aligned} \Omega_\mu &\equiv E\left[\left(\theta + c\left(\mu^\nabla\right) + \widetilde{\xi}_\mu\right)^2\right] - \left(\theta + c\left(\mu^\nabla\right)\right)^2 \\ &= E\left[\widetilde{\xi}_\mu^2\right] = \sigma_\mu^2, \end{aligned}$$

where $\widetilde{\xi}_\mu$ is the mean-zero variation in utility-space that (uniquely) implements μ and σ_μ^2 is the variance of $\widetilde{\xi}_\mu$. With $\sigma_\mu^2 = \beta^2\sigma_x^2$, we thus have

$$\begin{aligned} \Omega_\mu &= \left[c'\left(\mu^\nabla\right)\right]^2 \widehat{p}\,(1-\widehat{p}) \\ &\equiv \left[c'\left(\mu^\nabla\right)\right]^2 \widehat{\sigma}_x^2. \end{aligned}$$

independent of m as well. The (again, risk neutral) principal's problem then simply reduces to solving

$$\max_{\mu^\nabla} \quad \Pi\left(\mu^\nabla\right) - \left(\theta + c\left(\mu^\nabla\right)\right)^2 - \left[c'\left(\mu^\nabla\right)\right]^2 \widehat{\sigma}_x^2.$$

This implies that μ^{∇} solves

$$\frac{\Pi'\left(\mu^{\nabla}\right)}{2\left(\theta + c\left(\mu^{\nabla}\right) + c''\left(\mu^{\nabla}\right)\widehat{\sigma}_x^2\right)} = c'\left(\mu^{\nabla}\right), \tag{6}$$

and that, from (5),

$$\beta^{\nabla} = \frac{\Pi'\left(\mu^{\nabla}\right)}{2\kappa\left(\theta + c\left(\mu^{\nabla}\right) + c''\left(\mu^{\nabla}\right)\widehat{\sigma}_x^2\right)}$$

so that the fixed part of the agent's utility, $\alpha^{\nabla} \equiv \alpha\left(\mu^{\nabla}, \theta\right)$ hereafter, is determined as

$$\alpha^{\nabla} = \theta + c\left(\mu^{\nabla}\right) - \frac{\mu^{\nabla}\Pi'\left(\mu^{\nabla}\right)}{2\kappa\left(\theta + c\left(\mu^{\nabla}\right) + c''\left(\mu^{\nabla}\right)\widehat{\sigma}_x^2\right)}.$$

5 Applications

The point of developing this companion framework to Holmström and Milgrom (1987) is to, without loosing the indispensable tractability, provide slightly richer contracts that may be helpful in informing some of the discrepancies between the normative implications of the linear contracts and the empirical evidence accumulated. Furthermore, utilizing the approach of Hellwig and Schmidt (2002) to achieve the effort-independent variance result directly as the limiting case of the multinomial model adds additional structure that also has specific implications for the nature of the insights and predictions one can obtain here.

To facilitate specificity and tractability, as a base-line let $\Pi\left(\mu^{\nabla}\right) = \left(\mu^{\nabla}\right)^2 = \left(\kappa p^{\nabla}\right)^2$ and $c\left(\mu^{\nabla}\right) = \left(p^{\nabla}\right)^2/2$. Then, using (FOC) we have

$$\frac{\kappa^2 p^{\nabla}}{\theta + \left(p^{\nabla}\right)^2/2 + \widehat{\sigma}_x^2} = p^{\nabla}$$

or

$$p^{\nabla} = \sqrt{2\left(\kappa^2 - \theta - \widehat{\sigma}_x^2\right)} \tag{7}$$

so that

$$\mu^{\nabla} = \kappa\sqrt{2\left(\kappa^2 - \theta - \widehat{\sigma}_x^2\right)} \tag{8}$$

while

$$\beta^{\nabla} = \frac{\sqrt{2\left(\kappa^2 - \theta - \widehat{\sigma}_x^2\right)}}{\kappa} \tag{9}$$

and

$$\begin{aligned}
\alpha^{\nabla} &= \theta + p^2/2 - \beta\mu \\
&= \theta + \left(\kappa^2 - \theta - \widehat{\sigma}_x^2\right) - 2\left(\kappa^2 - \theta - \widehat{\sigma}_x^2\right) \\
&= 2\theta + \widehat{\sigma}_x^2 - \kappa^2.
\end{aligned} \tag{10}$$

While it might appear that the solution here is obtained from the standard Holmström (1979) formulation where the agent is held to his reservation utility, this is actually not the case for the following reason. While following Lemma 1 the second-best cost of implementing a particular μ, Ω_{μ}, *is* independent θ for the specific preference representation used here, the direct cost of compensating the agent for μ^{∇} is not - it is strictly increasing in θ due to the lack of a counterpart to Theorem 4 in Holmström and Milgrom (1987). Thus, while the principal's problem breaks down to maximizing the principal's residual net of second-best implementation costs and the first-best direct cost of procuring μ^{∇}, unlike in Holmström and Milgrom (1987) the specifics of the solution are not independent of how the economic "rents" are allocated. Accordingly, while the set-up has remained agnostic about the determinant(s) of θ so far, this issue has to be addressed before proceeding with comparative statics.

Regardless of how θ relates to the other parameters of the problem, with the specific functional forms chosen here the equilibrium value of the principal's objective function can be obtained as

$$\begin{aligned}\Gamma\left(\kappa,\theta,\widehat{\sigma}_x^2\right) &= \left(\kappa p^{\nabla}\right)^2 - \left(\theta^2+\theta\left(p^{\nabla}\right)^2+\left(p^{\nabla}\right)^4/4\right) - \left(p^{\nabla}\right)^2\widehat{\sigma}_x^2 \\ &= \left(p^{\nabla}\right)^2\left(\kappa^2-\theta-\widehat{\sigma}_x^2-\left(p^{\nabla}\right)^2/4\right)-\theta^2 \\ &= \left(p^{\nabla}\right)^2\left(\left(p^{\nabla}\right)^2/2-\left(p^{\nabla}\right)^2/4\right)-\theta^2 \\ &= \left(\kappa^2-\theta-\widehat{\sigma}_x^2\right)^2-\theta^2. \end{aligned} \tag{11}$$

Of course, if θ is independent of the specific parameters of the model, this would be consistent with θ representing the agent's best outside option in a competitive labor market. On the other end of the spectrum, however, principals may compete for agents to the point where the entire marginal product, somehow defined, accrues to the agent.

One straight-forward way to capture this is to assume that that the principal's acceptable expected (excess) net profit as represented by (11) is zero while the minimum exogenously determined expected utility acceptable to the agent is $\widehat{\theta} \geq 0$. Then, if the economic consequences of variations in κ^2, the marginal return to effort in this specification, accrue entirely to the agent, it follows from (11) that

$$\theta = \frac{\kappa^2-\widehat{\sigma}_x^2}{2}$$

so that

$$\frac{d\theta}{d\kappa^2} = -\frac{d\theta}{d\widehat{\sigma}^2} = \frac{1}{2}.$$

On the other hand, then, if the principal claims the entire surplus while the agent is held to his minimum expected utility it must be the case that

$$\widehat{\theta} \leq \frac{\kappa^2-\widehat{\sigma}_x^2}{2}$$

and independent of either κ^2 or $\widehat{\sigma}_x^2$. Finally, if the surplus is being shared so that both the principal and the agent earn positive "rents" we have $0 < d\theta/d\kappa^2 < 1/2$.

Because of the maintained ability to obtain closed form solutions to all endogenous variables, as summarized above, this slightly richer yet still relatively simple structure leads to a number of straight-forward but arguably more subtle comparative statics than the standard ones based off of Holmström and Milgrom (1987). Some specifics and suggestions are presented in the following sub-sections.

5.1 The "Risk" and PPS Relation

To identify the predicted relation between "risk" and pay-performance sensitivity (and thus "effort") for the model specification introduced above, consider the properties of the optimal cash compensation contract

$$s(X) = \alpha^2 + 2\alpha\beta X + \beta^2 X^2$$

Note first that PPS sensitivity,

$$\frac{ds(X)}{dX} = 2\alpha\beta + 2\beta^2 X$$

is increasing in X due to the convexity of the optimal contract. The expected (or average) PPS however is simply

$$E\left[\frac{ds(X)}{dX}\right] = 2\alpha\beta + 2\beta\mu\beta \tag{12}$$

which here depends on σ_x^2 only through its effect on μ since here α, β and μ are mechanically linked in equilibrium.

The right hand side of expression (12) is, of course, also the equivalent of the slope coefficient from a regression of compensation, $s(X)$, on performance X. To see this, let for expositional simplicity $A \equiv \alpha^2$, $B \equiv 2\alpha\beta$ and $C \equiv \beta^2$. The regression coefficients, φ_0 and φ_X from a standard OLS-regression of compensation on performance are then the solution to the following problem:

$$\min_{\varphi_0,\varphi_X} \quad E\left[\left(A+BX+CX^2-\varphi_0-\varphi_X X\right)^2\right]$$

which has first-order conditions

$$A-\varphi_0+(B-\varphi_X)\mu+C\left(\sigma^2+\mu^2\right)=0,$$
$$(A-\varphi_0)\mu+(B-\varphi_X)\left(\sigma^2+\mu^2\right)+C\left(3\mu\sigma^2+\mu^3\right)=0,$$

so that

$$\begin{aligned}\varphi_X &= 2C\mu+B \\ &= 2\alpha\beta+2\beta\mu\beta. \qquad (13)\end{aligned}$$

Thus, the predicted effect of risk on the slope coefficient from this regression is determined jointly by $\frac{d\beta}{d\sigma_x^2}$ and $\frac{d\alpha}{d\sigma_x^2}$. Note, however, that neither α or β depends directly on the variance of output, $\sigma_x^2 \equiv \kappa^2\widehat{\sigma}_x^2$; instead they depend of the components of σ_x^2. Further note, however, that κ^2 and $\widehat{\sigma}_x^2$ enter in the expressions for α, β and μ always with the opposite effect. Accordingly, *if* the PPS is estimated as the coefficient from an OLS-regression of compensation on performance, it may just as well increase as decrease in σ_x^2. It should thus come at no surprise either, that the substantial empirical literature that focus on this relation has been somewhat inconclusive.

The particular result that the PPS mcan just as well increase as decrease in the variance of the performance measure is not unique to my setting. Is due to the structure identified by Hellwig and Schmidt (2002) to ensure that performance variance is effort-independent in the continuos time limit and, thus, applies equally to the linear contract of Holmstrom and Milgrom (1987). This is of course somewhat ironic: under the specific conditions where the variance of the performance measure can be taken to be exogenous so that the comparative static for PPS w.r.t. σ_x^2 makes sense, this particular relation is indeterminate. This paradox does, however, help explain the very mixed evidence in the empirical literature on the relation between risk and PPS: because there is no theoretical prediction when variance is exogenous, there is no reason to believe the empirical evidence would be generally consistent with any particular hypothesis either.

While at first this conclusion may seem somewhat discouraging, it is, however, possible to empirically separate the portion of risk that captures fundamental variability, $\widehat{\sigma}_x^2$, from the κ^2-part that is proportional to the productivity (squared). The fact that σ_x^2 is relatively easily decomposed empirically actually provides for additional opportunities for better understanding the empirical properties of optimal compensation contract. The remainder of this section is dedicated to their issues of empirically decomposing the standard measure of performance measure variance into its two key component and then to use this decomposition to provide richer the standard PPS suggests new opportunities as the model does predict a negative relation between the fundamental risk component and *PPS* while a positive relation between the productivity component of risk and *PPS*. Specifically, if one was to look at stock returns over reasonable horizons, the proportion of positive returns is a direct estimate of $\widehat{p}_x$. With $\widehat{\sigma}_x^2 \equiv \widehat{p}_x \left(1 - \widehat{p}_x\right)$, κ^2 can then be directly recovered from the estimated overall variance of returns over the same horizon simply as $\sigma_x^2 / \widehat{\sigma}_x^2$.

Consider then first the extreme case where the Principal captures all the rents and $\theta \equiv \widehat{\theta}$ and therefore constant in both κ^2 and $\widehat{\sigma}_x^2$. Then substitute for the equilibrium values of α, β and μ in (13) to get

$$\widehat{\varphi}_X = 2\widehat{\theta}\frac{\sqrt{2\left(\kappa^2 - \widehat{\theta} - \widehat{\sigma}_x^2\right)}}{\kappa}$$

so that

$$\frac{d\widehat{\varphi}_X}{d\kappa} = 2\sqrt{2}\frac{\widehat{\theta}\left(\widehat{\theta} + \widehat{\sigma}_x^2\right)}{\kappa^2\sqrt{\kappa^2 - \widehat{\theta} - \widehat{\sigma}_x^2}}$$

and

$$\frac{d\widehat{\varphi}_X}{d\widehat{\sigma}_x^2} = -\sqrt{2}\frac{\widehat{\theta}}{\kappa\sqrt{\kappa^2 - \widehat{\theta} - \widehat{\sigma}_x^2}}.$$

The signs on the derrivatives are hardly surprising: PPS is increasing in the marginal product of "effort" and decreasing in "risk," at least when risk is properly defined as $\widehat{\sigma}_x^2$. What is, if not surprising then at least generally considered, is that the strength of the relations summarized by these derivatives depend on the allocation of "rents." To see this first note that both $\frac{d^2\widehat{\varphi}_X}{d\widehat{\theta}d\kappa}$ and $-\frac{d^2\widehat{\varphi}_X}{d\widehat{\theta}d\widehat{\sigma}_x^2}$ are positive.

Then consider the case where the agent captures all "rents." Using again (13), in this case

$$\varphi_X = \left(\kappa^2 - \widehat{\sigma}_x^2\right)\frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}}{\kappa}$$

so that

$$\frac{d\varphi_X}{d\kappa} = \frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}\left(2\kappa^2 + \widehat{\sigma}_x^2\right)}{\kappa^2}$$

and

$$\frac{d\varphi_X}{d\widehat{\sigma}_x^2} = -\frac{3}{2}\frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}}{\kappa}.$$

Then first compare φ_X to $\widehat{\varphi}_X$ for the extreme case of $\widehat{\theta} = \frac{\kappa^2 - \widehat{\sigma}_x^2}{2}$ or

$$\frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}\left(2\kappa^2 + \widehat{\sigma}_x^2\right)}{\kappa^2} \quad vs \quad 2\sqrt{2}\frac{\widehat{\theta}\left(\widehat{\theta} + \widehat{\sigma}_x^2\right)}{\kappa^2\sqrt{\kappa^2 - \widehat{\theta} - \widehat{\sigma}_x^2}}$$

$$\frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}\left(2\kappa^2 + \widehat{\sigma}_x^2\right)}{\kappa^2} \quad vs \quad 2\sqrt{2}\frac{\left(\frac{\kappa^2 - \widehat{\sigma}_x^2}{2}\right)^2 + \left(\frac{\kappa^2 - \widehat{\sigma}_x^2}{2}\right)\widehat{\sigma}_x^2}{\kappa^2\sqrt{\frac{\kappa^2 - \widehat{\sigma}_x^2}{2}}}$$

$$\frac{\sqrt{\kappa^2 - \widehat{\sigma}_x^2}\left(2\kappa^2 + \widehat{\sigma}_x^2\right)}{\kappa^2} \quad vs \quad \sqrt{2}\frac{\frac{\left(\kappa^2 - \widehat{\sigma}_x^2\right)^2}{2} + \left(\kappa^2 - \widehat{\sigma}_x^2\right)\widehat{\sigma}_x^2}{\kappa^2\sqrt{\kappa^2 - \widehat{\sigma}_x^2}}$$

$$\frac{\left(2\kappa^2+\widehat{\sigma}_x^2\right)}{\kappa^2} \qquad vs \qquad \sqrt{2}\frac{\frac{\left(\kappa^2-\widehat{\sigma}_x^2\right)^2}{2}+\left(\kappa^2-\widehat{\sigma}_x^2\right)\widehat{\sigma}_x^2}{\kappa^2\left(\kappa^2-\widehat{\sigma}_x^2\right)}$$

$$\left(2\kappa^2+\widehat{\sigma}_x^2\right) \qquad vs \qquad \left(\kappa^2-\widehat{\sigma}_x^2\right)+\widehat{\sigma}_x^2$$

$$\kappa^2+\widehat{\sigma}_x^2 \qquad vs \qquad 0$$

Accordingly, PPS increases in marginal pruduct and more so the more the agent's *expected* compensation increases in the marginal product.

Then do the same comparison for $\frac{d\widehat{\varphi}_X}{d\widehat{\sigma}_x^2}$ and $\frac{d\varphi_X}{d\widehat{\sigma}_x^2}$, *i.e.*,

$$-\frac{3}{2}\frac{\sqrt{\kappa^2-\widehat{\sigma}_x^2}}{\kappa} \qquad vs \qquad -\sqrt{2}\frac{\widehat{\theta}}{\kappa\sqrt{\kappa^2-\widehat{\theta}-\widehat{\sigma}_x^2}}$$

$$-\frac{3}{2}\frac{\sqrt{\kappa^2-\widehat{\sigma}_x^2}}{\kappa} \qquad vs \qquad -\frac{\sqrt{\kappa^2-\widehat{\sigma}_x^2}}{\kappa}$$

Accordingly, PPS decreases in fundamental risk and more so the more the agent's *expected* compensation increases in the marginal product.

5.2 Determinants of Convexity

An obvious issue to be investigated is the driver(s) of the optimal contract's "convexity." There are plenty of avenues to pursue but the particular route I'll take here is to look at the role of productivity, κ, in determining the weights on the linear and on the convex pieces, $2\alpha\beta$ and β^2 respectively. I'll start by considering the case where $\theta=\widehat{\theta}$ so that the principal captures all rents. To streamline this further, define $\overline{p}\equiv 1-\widehat{p}$, let $w=\overline{p}^2/2$ and $\kappa\in\left(\overline{p}^2/2+\widehat{\sigma}_x^2;\overline{p}^2+\widehat{\sigma}_x^2\right)$. The lower bound on κ result from $p\geq 0$ with the upper bound imposed by $p\leq 1$. While focusing on these bounds obviously is an extreme set-up with all the down sides that comprises, it has the advantage of delivering some crispness that could be considered helpful.

To see this note that at κ close to its lower bound here, β is (close to) zero while α is (close to) w. Moreover, we have

$$\frac{d\beta}{d\kappa} \to \infty$$

as κ approaches its lower bound. Accordingly, at the low end of the feasible productivity range, increased incentives in response to increased productivity are provided in the form of the linear component. The idea here would be that in (very) low-skill/productivity situations, incentives are provided in the from of piece rates: compensating based on the number of units produced, which does not seem ad odds with casual empiricism. Sharecroppers, newsboys, table servers all seem to fit with this.

On the other end of the productivity spectrum it is an entirely different story here, however. With the parametrization here, $\alpha \to 0$ and since

$$\frac{d\alpha}{d\kappa} < 0,$$

the weight on the linear component, $2\alpha\beta$, is strictly decreasing as κ increases towards its upper bound. Both β and $d\beta/d\kappa$ are, in contrast both strictly positive, so that at high levels of productivity, higher productivity implies substituting out of linear incentives and into convex ones. This seems, at least on the surface, consistent with options being more popular for top executives and more prevalent in strong growth environments such as the tech sector of the economy.

Moreover, there is an interaction-effect between the two pieces of overall risk in terms of the convexity of the optimal contract. Using the parameter values of the prior application, it is easily verified that for very high values of κ, $d\alpha\beta/d\widehat{\sigma}_x^2 > 0$ while $d\beta^2/d\widehat{\sigma}_x^2 < 0$. In contrast, $d\alpha\beta/d\widehat{\sigma}_x^2$ and $d\beta^2/d\widehat{\sigma}_x^2$ are both strictly negative for low values of κ. In other words, while increasing fundamental risk decreases (average) PPS regardless, in high productivity environments the linear component actually increases while at the low productivity side both risky components are scaled back.

Finally, then, consider the case where the agent gets all the rents. Note that in this case $\alpha = 0$, and all rewards to the agent then originates from the convex component so that the degree of convexity, as is PPS, is increasing in the marginal product and decreasing in the fundamental risk parameter. This suggests that high marginal product and a strong bargaining position for the agent both lead to more convex contracts. If the marginal product to some degree reflect managerial talent and if talent is in short supply it would seem likely that bargaining power and marginal product should be positively related. This, in turn, should then strengthen the proposed positive relation between convexity and marginal product identified in the case where the principal keeps all the rents.

5.3 Performance Measure Properties and Design

The final implication pursued here is that of the design of an optimal performance measure and the empirical fingerprint associated with better performance measures. What is of particular interest in this section is the dynamic properties of the aggregate performance measure X. The main insight is again subtle but important: while the optimal contract can be written just on aggregate performance, the *parameters* and the *efficiency* of the contract depends critically on the time-series properties of measured performance holding all else constant. That is, performance measures that are indistinguishable on all dimensions in the aggregate but do not possess the same dynamic properties are not equal from the perspective of the second-best. Aggregates are sufficient, but aggregates are not sufficient for evaluating how good they actually are from the perspective of contracting efficiency.

For expositional ease, consider the semi-parametric mode introduced above as a benchmark with a bit of additional structure to the specifics of the performance measure, x. Specifically, introduce a separate productivity parameter, say γ, so that the probability of x_1^Δ, $Pr\left(x_1^\Delta|\widehat{p}, p, \gamma\right) \equiv p_\gamma^\Delta = \widehat{p} + \Delta^{1/2}\gamma\left(p - \widehat{p}\right)$. Then consider any class of performance measures, $\overline{X}$, for which i) the variance $\kappa^2\widehat{p}\left(1-\widehat{p}\right) = \underline{\sigma}^2$, ii) the sensitivity to the agent's action $d\mu^\Delta/dp^\Delta = \underline{\delta}$, and the mean is the same given the agent's action which, given ii), is implied by $iii)$ $\widehat{\mu} = \underline{\widehat{\mu}}$. Notice that $i)$ implies

$$\kappa = \sqrt{\frac{\underline{\sigma}_x^2}{\widehat{p}(1-\widehat{p})}},$$

$ii)$ implies

$$\gamma = \underline{\delta}/\kappa,$$

while $iii)$ implies

$$x_0 = \underline{\widehat{\mu}} - \widehat{p}\kappa.$$

In other words while $i)$ imposes an implicit relation between κ and $\widehat{p}$, the free parameters γ and x_0 can always be adjusted to satisfy $ii)$ and $iii)$. Importantly, it is easily verified that as a result of this, changing $\widehat{p}$ while satisfying $i)$, $ii)$ and $iii)$ does not affect the structure of the solution to the principal's problem as represented by (7)-(10) here. What does have to change, however, is $\widehat{\sigma}_x^2$ which is maximized for $\widehat{p} = .5$ and is approaching zero as $|\widehat{p} - .5| \rightarrow .5$. This is important for two reasons: first while $\widehat{\sigma}_x^2$ is a fundamental driver of contracting efficiency variations in $\widehat{\sigma}_x^2$ is not easily identified identified by the properties aggregate performance. Where $\widehat{\sigma}_x^2$ *is* more easily detected is in the time-series properties of performance: if performance evolves symmetrically around its mean $\widehat{\sigma}_x^2$ is (close to being) at its max. On the other hand a measure with more frequent smaller gains and fewer but bigger losses or vice versa, fewer but larger gains but more frequent smaller losses, indicate lower values of $\widehat{\sigma}_x^2$.

This seems particularly interesting from an accounting perspective. Accounting conventions are broadly about creating timing differences and thus about altering the time-series properties of earnings. Under something akin to clean surplus accounting the properties of *aggregate* earnings over a reasonable horizon are however fundamentally unaffected by the particular approach to accounting measurement chosen by the firm. In terms of the model, this is exactly the consequence of being able to choose a $\widehat{p}$ without being able to alter the *aggregate* mean, variance or sensitivity to effort over a reasonable horizon. Moreover, choices of $\widehat{p} \neq .5$ have reasonably straight forward interpretation in terms of accounting measurement bias: few "big" write-offs in conjunction with being able to produce smaller positive surprises over longer horizons (think "big bath" behavior in the extreme) is generally considered a feature of aggressive reporting while making overly cautious loss provisions for many periods that on average then all reverse in the period of repayment is generally considered conservative accounting. In the model, either is strictly preferred to the neutral case.

In addition to thus suggesting value of biased accounting measures it also offer a means to reconcile the apparent appeal of both conservative and aggressive practices: timing induced bias, regardless of the direction of the bias, reduces the fundamental variance of the performance measure and improves contracting efficiency. This, of course also brings about some suggestions for empirical inquiry along the lines of what was discussed in the previous sub-section. While γ generally is unobservable, holding aggregate (average) performance and performance variance constant, the degree of asymmetries in periodic accounting performance continues to provide a measure of $\widehat{\sigma}_x^2$. This, in turn, provide a link between the time-series properties of measured performance, the convexity of contracts and their average PPS. Particularly, it predicts that there is not a monotone relation between accounting bias and properties of accounting based contracts: higher biases leads to higher average PPS. While obviously only very simple and incomplete, this at least suggest that the framework here proposed, may have some promise in generating specific easily testable predictions along a number of such dimensions.

6 Conclusion

In this paper I develop a simple agency framework based on the dynamic approach first proposed by Holmström and Milgrom (1987). The key distinguishing feature here is that I rely on an additively separable (in aggregate consumption and periodic effort) preference structure instead of the multiplicatively separable negative exponential representation in their original specification. I proceed to show that when the agent has square-root preferences for aggregate consumption and the risk-neutral principal must choose a production technology evolution for the contracting horizon to match future action choices by the agent at the start of this horizon, the key stationarity result of Holmström and Milgrom (1987) re-emerges in my setting. Combined with the uniqueness of the contract that implements a given action in each sub-period, their main aggregation result applies here as well: the optimal contract can be written as a simple linear function of the aggregate balances of a set of enumeration accounts. The key difference here is that while the contract in the set-up in Holmström and Milgrom (1987) is linear in cash-space, with the additively separable specification I use it is linear in utility-space and, thus, adding a convex component to the linear in cash-space.

Relying on Hellwig and Schmidt (2002) to obtain the continuos time Brownian motion version from the discrete time binomial version of the model as the limiting case when the length of the sub periods approach zero, provides additional structure that when paired with some additional simplifying (standard) assumptions on the nature of the production and cost functions along with on how actions are captured makes the framework as amenable to comparative statics as the so called LEN model that originated in Holmström and Milgrom (1987). In fact, because my specification includes an easily identifiable convex component and additional structure on the components of performance volatility, the set of comparative statics and directly testable empirical predictions that can be obtained here holds real promise for enhancing our understanding of the determinants of optimal contracts as well as optimal performance measure design. The paper closes by providing a number of examples suggestive of this potential.

References

[1] Edmans, A., and X. Gabaix, 2011, Tractability in Incentive Contracting. *The Review of Financial Studies* 24, 2865-94

[2] Grossman, S., and O. Hart, 1983, An Analysis of the Principal-Agent Problem. *Econometrica* 51, 7-45.

[3] Hellwig, M., and K. Schmidt, 2002, Discrete-Time Approximation of the Holmström-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision. *Econometrica* 70, 2225-64.

[4] Hemmer, T., 2017, Income Smoothing as Rational Equilibrium Behavior? A Second Look. *Working Paper.*

[5] Holmström, B., and P. Milgrom, 1987, Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55, 303-28.