Tractability in Incentive Contracting

Alex Edmans
Wharton School, University of Pennsylvania and NBER

Xavier Gabaix
NYU Stern School of Business, CEPR and NBER

This article develops a framework that delivers tractable (i.e., closed-form) optimal contracts, with few restrictions on the utility function, cost of effort, or noise distribution. By modeling the noise before the action in each period, we force the contract to provide correct incentives state-by-state, rather than merely on average. This tightly constrains the set of admissible contracts and allows for a simple solution to the contracting problem. Our results continue to hold in continuous time, where noise and actions are simultaneous. We illustrate the potential usefulness of our setup by a series of examples related to CEO incentives. In particular, the model derives predictions for the optimal measure of incentives and whether the contract should be convex, concave, or linear. (*JEL* D86, G34)

The principal-agent problem is central to many settings in economics and finance, such as compensation, insurance, taxation, and regulation. A vast literature analyzing this problem has found that it is typically difficult to solve, even in simple settings. The first-order approach is often invalid, requiring the use of more intricate techniques. Even if an optimal contract can be derived, it is often not attainable in closed form, which reduces tractability—a particularly important feature in applied theory models.

This article develops a broad framework that delivers tractable, closed-form contracts, with few restrictions on the utility function, cost of effort, or noise distribution. The framework requires two conditions: the analysis of a given path of effort levels, and either continuous time or a discrete-time model with a modified timing assumption. Grossman and Hart (1983) show that the

discrete-time problem is complex even when the target action is fixed, since many contracts can implement the target action. We achieve tractability by specifying that, in each period, the agent exerts effort after observing the noise (and then observes the noise in the next period). This is similar to theories in which the agent observes total cash flow before deciding how much to divert (e.g., Lacker and Weinberg 1989; Biais et al. 2007; DeMarzo and Fishman 2007). Since the agent knows the noise when taking his action, incentive compatibility requires the agent's marginal incentives to be correct state-by-state (i.e., for every possible noise outcome), which tightly constrains the set of admissible contracts. By contrast, if noise followed the action, incentive compatibility would only pin down incentives in expectation. Many contracts provide correct incentives on average, and the problem is complex as the principal must solve for the cheapest contract out of this continuum. Note that the timing assumption does not change the fact that the agent faces uncertainty when deciding his effort since each action, except the final one, is still followed by noise. Even in a one-period model, the agent faces risk as the noise is unknown when he signs the contract. We then show that the contract retains the same form in continuous time where noise and effort occur simultaneously. This consistency suggests that, if underlying reality is continuous time, it is best approximated in discrete time by modeling noise before effort in each period.

Tractability allows the economic forces driving the contract to be transparent; in particular, we can see what features of the environment do and do not matter for the contract. Its functional form is independent of the agent's noise distribution and reservation utility, and depends only on how the agent trades off the benefits of cash against the cost of providing effort. Moreover, the contract's slope, as well as its functional form, is independent of the agent's utility function, reservation utility, and noise distribution in two cases. First, if the cost of effort is pecuniary (i.e., can be expressed as a subtraction to cash pay), the incentive scheme is linear in output regardless of these parameters. Second, if the agent's preferences are multiplicative in cash and effort, the contract is independent of these parameters and log-linear. This robustness contrasts many classical principal-agent models (e.g., Grossman and Hart 1983), where the contract is contingent upon many specific features of the setting. Our results imply that, under some specifications, the contract is robust to such parametric uncertainty.

We next allow the target effort level to depend on the current-period noise, similar to papers in which the agent observes the state of nature before choosing his action.[1] The principal now implements an "action function," which specifies a different action for each noise realization. We identify the class of feasible action functions, providing a necessary and sufficient condition for

---

[1] See, for example, Harris and Raviv (1979), Sappington (1983), Baker (1992), and Prendergast (2002). In Laffont and Tirole (1986), the firm observes its efficiency before taking its actions. However, in that paper, the efficiency is known before contracting (it is a standard adverse selection model); here, and in the above papers, the noise/ state of nature is revealed after contracting but before the action.

a given action function to be implementable. The contract now depends on messages sent by the agent regarding the noise, but remains tractable.

The above analysis focuses on the implementation of a given action function. Jointly deriving the optimal action in addition to the efficient contract that implements it is typically extremely complex. Studying a given effort level allows for significant tractability and is useful for practical applications. Consequently, many contracting papers focus exclusively (e.g., Dittmann and Maug 2007; Dittmann, Maug, and Spalt 2010) or predominantly (e.g., Grossman and Hart 1983; Lacker and Weinberg 1989) on implementing a given effort level.

We derive a sufficient condition under which the optimal action (out of a continuous action space) is independent of the current noise, and thus it is sufficient to focus on a fixed action to solve the full contracting problem. The efficient action is a tradeoff between the benefits and costs of effort. The former are of similar order of magnitude to the output under the agent's control, and the latter (disutility plus the risk imposed by incentives) are of similar order of magnitude to the agent's wage. Thus, if output is sufficiently large (e.g., the agent is a CEO who controls a firm), the benefits of effort swamp the costs. Therefore, inducing the highest productive level of effort is optimal regardless of the noise (the "high effort principle"). In a cash flow diversion model, full productive efficiency corresponds to zero stealing; in a project selection model, it corresponds to taking all positive-NPV projects while rejecting negative-NPV ones. The analysis thus demonstrates the conditions under which it is justifiable to focus on a fixed effort level, such as for CEOs or other agents with a large effect on outcomes.

Finally, we allow the principal to choose the highest productive effort level by extending the model to a two-stage game. In the first stage, the principal makes an irreversible choice of productive capacity (e.g., by selecting plant size), that determines the highest productive effort level. In the second stage, the contract is played out as before—the principal wishes the agent to run the plant (whatever its size) with full efficiency. As in standard models, the effort level set in the first stage is typically decreasing in disutility, risk aversion, and risk. Thus, the two-stage game allows for contracts that are simple (since high effort is optimal in the second stage and so solving a complex tradeoff to derive the optimal effort level is not required) yet still respond to the features of the setting and thus generate comparative statics.

In sum, our analysis generates a set of sufficient conditions to obtain tractable contracts: ex-post actions plus a high benefit of effort. (Holmstrom and Milgrom 1987, "HM") developed a different set of sufficient conditions under which the optimal contract is tractable (indeed linear): exponential utility, a pecuniary cost of effort, Gaussian noise, and continuous time. Their result has since been widely used by applied theorists to justify assuming linear contracts (e.g., Baranchuk, Macdonald, and Yang 2011). However, the required conditions may not hold in a number of situations. For example, power utility

is often used given evidence of decreasing absolute risk aversion; many agent actions do not involve a monetary expenditure. Perhaps because some of the above conditions are not met, contracts are not always linear in reality—for instance, managerial contracts are often convex. Our framework develops a quite different set of sufficient conditions, which may be satisfied in many settings in which the HM assumptions do not hold and tractability was previously believed to be unattainable. In addition, while the HM setup delivers linear contracts, our setting also accommodates convex and concave contracts.

An application to CEO incentives demonstrates the additional implications that can be obtained by allowing for general utility and cost functions. One relates to the optimal measure of incentives. For CEOs, Edmans, Gabaix, and Landier (2009) show that multiplicative preferences are necessary to obtain empirically consistent predictions. The contract is thus log-linear in performance; since the appropriate output measure is the stock return, the contract relates the percentage change in pay to the percentage change in firm value. This analysis provides a theoretical justification for measuring incentives using the elasticity of pay to firm value, a metric previously advocated by Murphy (1999) on empirical grounds. A second implication is on the structure of compensation. In practice, CEOs are paid with options as well as stock, leading to convex contracts. Since standard models with exponential utility predict concave contracts (Dittmann and Maug 2007), some commentators argue that the use of options is inefficient (Bebchuk and Fried 2004). In standard models, whether a contract is concave or convex is often difficult to determine analytically and so calibration must be used (Dittmann and Maug 2007). Our closed-form solutions show that the contract's shape depends only on the marginal cost of effort; if it exceeds unity, the contract is convex, as in reality. Third, since our framework does not require exponential utility, it allows for wealth effects and thus additional comparative statics. Edmans and Gabaix (2011) use this framework to show that, in a market equilibrium, these wealth effects cause distortions in CEO assignment and aggregate production. Finally, the tractability of the framework allows it to be easily extended to accommodate other agent actions, such as intermediate consumption, private saving, and short-termism, while retaining realistic wealth effects (Edmans et al. 2011).

In addition to its results, the article's proofs import and extend some mathematical techniques that are rare in economics and may be of use in future models. We use the subderivative[2] to avoid the first-order approach, and so it may be useful for models where sufficient conditions for the first-order approach cannot be verified. We also use "relative dispersion" to prove that the incentive constraints bind (i.e., the principal imposes the minimum slope that induces effort), and also to rule out stochastic contracts, where the payout is a random function of output.

---

[2] This is a generalization of the derivative that allows for quasi first-order conditions even if the objective function is not everywhere differentiable.

This article builds on a rich literature on tractable multiperiod agency problems. Mueller (2000) and Hellwig and Schmidt (2002) study the extent to which the HM framework can be extended to discrete time, and Sung (1995) and Ou-Yang (2003) allow the agent to control the diffusion of returns as well as the drift. All of these papers require exponential utility and a pecuniary cost of effort. Our modeling of noise before the action is most similar to models in which the agent can observe total cash flow before deciding how much to divert. Lacker and Weinberg (1989) show that the optimal contract to deter all diversion (the analog of highest effort) is piecewise linear, regardless of the noise distribution and utility function. Their core result is similar to a specific case of our Theorem 1, restricted to a pecuniary cost of effort and a single period. DeMarzo and Sannikov (2006), Biais et al. (2007), and DeMarzo and Fishman assume risk neutrality, and so the contract is linear.

This article proceeds as follows. In Section 1, we derive tractable contracts under a constant implemented action, in both discrete and continuous time. Section 2 allows for the target action to depend on the realized noise, derives a condition under which the optimal action is deterministic (independent of the noise), and allows the principal to choose the optimal action according to the parameters of the environment. Section 3 concludes. The Appendix contains proofs not in the main text, along with additional peripheral material.[3]

1. Optimal Contract with Deterministic Actions

1.1 Discrete time

We consider a T-period model; its key parameters are summarized in Table 1. In each period t, the agent observes noise η_t, takes an unobservable action a_t, and then observes the noise in period $t + 1$. The action a_t is broadly defined to encompass any decision that benefits output but is costly to the agent. Examples include effort (low a_t represents shirking), project choice (low a_t involves selecting projects that maximize private benefits rather than firm value), or rent extraction (low a_t reflects cash flow diversion). Noises $\eta_1, ..., \eta_T$ are independent with interval support with interior $\left(\underline{\eta_t}, \overline{\eta_t} \right)$, where the bounds may be infinite, and that $\eta_2, ..., \eta_t$ have log-concave densities.[4] We require no other distributional assumption for η_t; in particular, it need not be Gaussian. The action space $\mathcal{A}$ has interval support, bounded below and above by $\underline{a}$ and $\overline{a}$. We allow for both open and closed action sets and for the bounds to be infinite. After the action is taken, a verifiable signal,

$$r_t = a_t + \eta_t, \tag{1}$$

is publicly observed at the end of each period t.

---

[3] All appendices are available online at http://www.sfsrfs.org.

[4] A random variable is log-concave if it has a density with respect to the Lebesgue measure, and the log of this density is a concave function. Many standard density functions are log-concave, in particular the Gaussian, uniform, exponential, Laplace, Dirichlet, Weibull, and beta distributions (e.g., Caplin and Nalebuff 1991). On the other hand, most fat-tailed distributions are not log-concave, such as the Pareto distribution.

Table 1
Key variables in the model

Variable	Description
a	Effort (also referred to as "action")
$\overline{a}$	Highest effort
$\overline{\overline{a}}$	Highest productive effort
a^*	Target effort
b	Benefit function for effort, defined over a
c	Cash compensation, defined over r or η
f	Density of the noise distribution
g	Cost of effort, defined over a
r	Signal (or "return"), typically $r = a + \eta$
u	Agent's utility function, defined over $v(c) - g(a)$
$\underline{u}$	Agent's reservation utility
v	Agent's felicity function, defined over c
η	Noise
A	Action function, defined over η
$C[A]$	Expected cost of contract implementing $\left\{ A(\eta), \eta \in \left(\underline{\eta}, \overline{\eta} \right) \right\}$
$\overline{F}$	Complementary cumulative distribution function of η
S	Baseline size of output under agent's control
T	Number of periods
V	Felicity provided by contract, defined over r or η

In period T, the principal pays the agent cash of c.[5] The agent's utility function is

$$E\left[u\left(v(c) - \sum_{t=1}^{T} g(a_t) \right) \right]. \qquad (2)$$

g represents the cost of effort, which is increasing and weakly convex. u is the utility function, and v is the felicity[6] function that denotes the agent's utility from cash; both are increasing and weakly concave. g, u, and v are all twice continuously differentiable. We specify functions for both utility and felicity to maximize the generality of the setup. For example, the utility function $\left(ce^{-g(a)} \right)^{1-\gamma} / (1 - \gamma)$ is often used in macroeconomics (e.g., Cooley and Prescott 1995) and in the executive compensation models of Edmans and Gabaix (2011) and Edmans et al. (2011) to obtain realistic income effects, which entails $u(x) = e^{(1-\gamma)x} / (1 - \gamma)$ and $v(x) = \ln x$. The case $u(x) = x$ denotes additively separable preferences; $v(c) = \ln c$ generates multiplicative preferences. If $v(c) = c$, the cost of effort is expressed as a subtraction to cash pay. This is appropriate if effort represents an opportunity cost of foregoing an alternative income-generating activity, or involves a financial expenditure; however, most effort decisions (e.g., foregoing leisure or private benefits) do not involve a pecuniary cost. HM assume $u(x) = -e^{-\gamma x}$ and $v(c) = c$. The only assumption that we make for u is that it exhibits nonincreasing absolute

---

[5] If the agent quits before time T, he receives a very low wage $\underline{c}$.

[6] We note that the term "felicity" is typically used to denote one-period utility in an intertemporal model. We use it in a non-standard manner here to distinguish it from the utility function u.

risk aversion (NIARA), i.e., $-u''(x)/u'(x)$ is nonincreasing in x. Most common utility functions (e.g., constant absolute risk aversion $u(x) = -e^{-\gamma x}$ and constant relative risk aversion $u(x) = x^{1-\gamma}/(1-\gamma)$, $\gamma > 0$) exhibit NIARA. This assumption is sufficient to rule out randomized contracts.

The agent's reservation utility is $\underline{u} \in \operatorname{Im} u$, where $\operatorname{Im} u$ is the image of u, i.e., the range of values taken by u. We assume that $\operatorname{Im} v = \mathbb{R}$ so that we can apply the v^{-1} function to any real number.[7] We take an optimal contracting approach that imposes no restrictions on the contracting space available to the principal, so the contract $\widetilde{c}(\cdot)$ can be stochastic and nonlinear in the signals r_t. The timing is as follows:

1. The principal proposes a (possibly stochastic) contract $\widetilde{c}(r_1, ..., r_T)$.
2. The agent agrees to the contract or receives his reservation utility $\underline{u}$.
3. The agent observes noise η_1, then exerts effort a_1.
4. The signal $r_1 = \eta_1 + a_1$ is publicly observed.
5. Steps (3)–(4) are repeated for $t = 2, ..., T$.
6. The principal pays the agent $\widetilde{c}(r_1, ..., r_T)$.

Throughout most of the article, we abstract from imperfect commitment problems and focus on a single source of market imperfection: moral hazard. This assumption is common in the dynamic moral hazard literature (e.g., Rogerson 1985, HM, Spear and Srivastava 1987, Phelan and Townsend 1991, Biais et al. 2007, Biais et al. 2010). Appendix E extends the model to accommodate quits and firings.

As in the first stage of Grossman and Hart (1983), we initially fix the path of effort levels that the principal wants to implement at $(a_t^*)_{t=1,..,T}$, where $a_t^* > \underline{a}$ and a_t^* may be time-varying.[8] In Section 2.1, we allow for the target action to depend on the current noise. An admissible contract gives the agent an expected utility of at least $\underline{u}$ and induces him to take path (a_t^*) and truthfully report noises $(\eta_t)_{t=1,..,T}$. The principal is risk-neutral, and so the optimal contract is the admissible contract with the lowest expected cost $\mathbb{E}[\widetilde{c}]$. In Appendix B, we allow for the contract to depend on messages sent by the agent on the noise η_t, and show that they are redundant. Intuitively, since a_t^* is implemented for all η_t, there is a one-to-one correspondence between r_t and η_t on the equilibrium path. The principal can thus infer η_t from r_t, rendering messages redundant.

We now formally define the principal's program. Let $\mathcal{F}_t$ be the filtration induced by $(\eta_1, ..., \eta_t)$, the noise revealed up to time t. The agent's policy is $(a) = (a_1, ..., a_T)$, where a_t is the effort taken if noise $(\eta_1, ..., \eta_t)$ has been

---

[7] This assumption could be weakened. With K defined as in Theorem 1, it is sufficient to assume that there exists a value of K that makes the participation constraint bind, and a "threat consumption," which deters the agent from exerting very low effort, i.e., $\inf_c v(c) - \inf_{a_t} \sum_t g(a_t) \leq \sum_t g'(a^*)\left(\underline{\eta_t} + a_t^*\right) + K$.

[8] If $a_t^* = \underline{a}$, then a flat wage induces the optimal action.

realized, and is $\mathcal{F}_t$–measurable. Define $(a^*) = (a_1^*, ..., a_T^*)$ as the policy of exerting effort a_t^* at time t. The program is given below:

Program 1.   Assume $a_t^* > \underline{a} \ \forall \ t$. The principal chooses a contract $\widetilde{c}(r_1, ..., r_T)$ that minimizes expected cost:

$$\min_{\widetilde{c}(\cdot)} \ \mathrm{E}\left[\widetilde{c}\left(a_1^* + \eta_1, ..., a_T^* + \eta_T\right)\right], \tag{3}$$

subject to the following constraints:

$$\forall \eta_1, ..., \eta_t, \ a_t^* \in \arg\max_{a_t} \mathrm{E}\left[u\left(v\left(\widetilde{c}\left(a_1^* + \eta_1, ..., a_t + \eta_t, ..., a_T^* + \eta_T\right)\right)\right.\right.$$

$$\left.\left. -g\left(a_t\right) - \sum_{s=1,s\neq t}^{T} g\left(a_s^*\right)\right) \mid \eta_1, ..., \eta_t\right]. \tag{4}$$

$$\mathrm{IR:} \ \mathrm{E}\left[u\left(v\left(\widetilde{c}\left(\cdot\right)\right) - \sum_{t=1}^{T} g\left(a_t^*\right)\right)\right] \geq \underline{u}. \tag{5}$$

Theorem 1 describes our solution to Program 1.[9]

Theorem 1.   (Optimal Contract, Discrete Time) *The following contract is optimal. The agent is paid*

$$c = v^{-1}\left(\sum_{t=1}^{T} g'\left(a_t^*\right) r_t + K\right), \tag{6}$$

where K is a constant that makes the participation constraint bind $\left(\mathrm{E}\left[u\left(\begin{array}{c}\sum_t g'\left(a_t^*\right) r_t + \\ K - \sum_t g\left(a_t^*\right)\end{array}\right)\right] = \underline{u}\right)$. The functional form (6) is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η; these parameters affect only the scalar K. The optimal contract is deterministic.

In particular, if the target action is time-independent ($a_t^* = a^* \ \forall \ t$), the contract

$$c = v^{-1}\left(g'\left(a^*\right) r + K\right) \tag{7}$$

is optimal, where $r = \sum_{t=1}^{T} r_t$ is the total signal.

---

[9] Theorem 1 characterizes a contract that is optimal, i.e., solves Program 1. Strictly speaking, there exist other optimal contracts that pay the same as (6) on the equilibrium path, but take different values for returns that are not observed on the equilibrium path. Note that the contract in Theorem 1 allows c to be negative. Limited liability could be incorporated, at the cost of additional notational complexity, by imposing a lower bound on $\underline{\eta}$ or adding a fixed constant to the signal.

Proof. (Heuristic). Appendix B presents a rigorous proof that rules out stochastic contracts, and does not assume that the contract is differentiable. Here, we give a heuristic proof by induction on T that conveys the essence of the result for deterministic contracts, using first-order conditions and assuming $a_t^* < \bar{a}$. We commence with $T = 1$. Since η_1 is known, we can remove the expectations operator from the IC condition (4). Since u is an increasing function, it also drops out to yield

$$a_1^* \in \arg\max_{a_1} v\left(c\left(a_1 + \eta_1\right)\right) - g\left(a_1\right). \tag{8}$$

The first-order condition is

$$v'\left(c\left(a_1^* + \eta_1\right)\right) c'(a_1^* + \eta_1) - g'\left(a_1^*\right) = 0. \tag{9}$$

Therefore, for all r_1,

$$v'\left(c\left(r_1\right)\right) c'\left(r_1\right) = g'\left(a_1^*\right),$$

which integrates over η_1 to

$$v\left(c\left(r_1\right)\right) = g'\left(a_1^*\right) r_1 + K \tag{10}$$

for some constant K. Contract (10) must hold for all r_1 that occurs with non-zero probability, i.e., for $r_1 \in \left(a_1^* + \underline{\eta}_1, a_1^* + \overline{\eta}_1\right)$.

We will proceed now by induction on the total number of periods T: We now show that, if the result holds for T, it also holds for $T+1$. Let $V\left(r_1, ..., r_{T+1}\right) \equiv v\left(c\left(r_1, ..., r_{T+1}\right)\right)$ denote the indirect felicity function, i.e., the contract in terms of felicity rather than cash. At $t = T + 1$, the IC condition is

$$a_{T+1}^* \in \arg\max_{a_{T+1}} V\left(r_1, ..., r_T, \eta_{T+1} + a_{T+1}\right) - g\left(a_{T+1}\right) - \sum_{t=1}^{T} g\left(a_t^*\right). \tag{11}$$

Applying the result for $T = 1$, to induce a_{T+1}^* at $T + 1$, the contract must be of the form

$$V\left(r_1, ..., r_T, r_{T+1}\right) = g'\left(a_{T+1}^*\right) r_{T+1} + k\left(r_1, ..., r_T\right), \tag{12}$$

where the integration "constant" now depends on the past signals, i.e., $k(r_1, ..., r_T)$. In turn, $k\left(r_1, ..., r_T\right)$ is chosen to implement $a_1^*, ..., a_T^*$ viewed from $t = 0$, when the agent's utility is

$$E\left[u\left(k\left(r_1, ..., r_T\right) + g'\left(a_{T+1}^*\right) r_{T+1} - g\left(a_{T+1}^*\right) - \sum_{t=1}^{T} g\left(a_t\right)\right)\right].$$

Defining

$$\widehat{u}\left(x\right) = E\left[u\left(x + g'\left(a_{T+1}^*\right) r_{T+1} - g\left(a_{T+1}^*\right)\right)\right], \tag{13}$$

the principal's problem is to implement $a_1^*, ..., a_T^*$ with a contract $k(r_1, ..., r_T)$, given a utility function

$$E\left[\widehat{u}\left(k(r_1, ..., r_T) - \sum_{t=1}^{T} g(a_t)\right)\right].$$

Applying the result for T, the contract must have the form $k(r_1, ..., r_T) = \sum_{t=1}^{T} g'(a_t^*) r_t + K$ for some constant K. Combining this with (10), the contract must satisfy

$$V(r_1, ..., r_T, r_{T+1}) = \sum_{t=1}^{T+1} g'(a_t^*) r_t + K, \tag{14}$$

for (r_t) that occurs with non-zero probability (i.e., $(r_1, ..., r_T) \in \prod_{t=1}^{T}\left(a_t^* + \underline{\eta}_t, a_t^* + \overline{\eta}_t\right)$). The associated pay is $c = v^{-1}\left(\sum_{t=1}^{T+1} g'(a_t^*) r_t + K\right)$, as in (6). Conversely, any contract that satisfies (14) is incentive compatible. ∎

Theorem 1 yields a closed-form contract for any T and (a_t^*). It also clarifies the parameters that do and do not matter for the contract's functional form. It depends only on the felicity function v and the cost of effort g, i.e., how the agent trades off the benefits of cash against the costs of providing effort. It is independent of the utility function u, the reservation utility $\underline{u}$, and the distribution of the noise η, i.e., can be written without reference to these parameters. Even though these parameters do not affect the contract's functional form, in general they will affect its slope via their impact on the scalar K. However, if $v(c) = c$ (the cost of effort is pecuniary), the contract's slope is also independent of u, $\underline{u}$, and η: It is linear, regardless of these parameters. The linear contracts of HM can thus be achieved in settings that do not require exponential utility, Gaussian noise, or continuous time. (Note that, even if the cost of effort is pecuniary, it remains a general, possibly nonlinear function $g(a_t)$.) If $v(c) = \ln c$, the contract's slope is also independent of u, $\underline{u}$, and the distribution of η. This "detail-independence" contrasts with standard agency models where the contract depends on many specific features of the setting. This poses practical difficulties, as some of the important determinants are difficult for the principal to observe and thus use to guide the contract, such as the noise distribution and utility function. Our results provide a taxonomy of situations in which the contract is robust to parametric uncertainty.[10] The framework thus offers a potential explanation for why real-world contracts do not seem to be as complicated and contingent on as many details of the environment as standard contract theories would suggest.

---

[10] Chassang (2011) also derives detail-independent contracts in a different setting.

The intuition for why "noise-before-action" timing allows us to dispense with the HM assumptions, yet still achieve tractability, can be seen in the heuristic proof. We first consider $T = 1$. Since η_1 is known, the expectations operator can be removed from (4). u then drops out to yield (8). The specific form of u is irrelevant—all that matters is that it is monotonic, and so it is maximized by maximizing its argument. In particular, the HM assumption of exponential utility is not required—the agent's attitude to risk does not matter, as η_1 is known. In turn, (8) yields the first-order condition (9), which must hold for every possible realization of η_1, i.e., *state-by-state*. This pins down the slope of the contract: For all η_1, the agent must receive a marginal felicity of $g'\left(a_1^*\right)$ for a one-unit increment to the signal r_1. The principal's only degree of freedom is the constant K, which is itself pinned down by the participation constraint.

By contrast, if η_1 followed the action, and assuming linear u for simplicity, (9) would be

$$E\left[v'\left(c\left(r_1\right)\right)c'\left(r_1\right)\right] = g'\left(a_1^*\right). \tag{15}$$

This first-order condition only determines the agent's marginal incentives *on average*, rather than state-by-state. Multiple contracts satisfy (9) and implement a_1^*, and so the problem is highly complex (even in a single period) as the principal must solve for the cheapest contract out of this continuum. HM tighten the set of admissible contracts by giving the agent substantial freedom in two ways: He controls the probabilities of N different states of nature and chooses his actions in continuous time. We instead give the agent freedom by specifying the noise before the action, which allows us to dispense with continuous time and model the action more simply as the choice of the mean return, as is common in applied theory models.

Even though all noise is known when the agent takes his action, it is not automatically irrelevant. First, since the agent does not know η_1 when he signs the contract, he is subject to risk and so the first-best is not achieved (see Edmans and Gabaix (2011) for the distortions this leads to in a market equilibrium). Second, the noise realization has the potential to undo incentives. If η_1 is high, r_1 and thus c will already be high; a high $\underline{u}$ has the same effect. If the agent exhibits diminishing marginal felicity (i.e., v is concave), he has lower incentives to exert effort. Put differently, when the agent takes his action, he does not face *risk* (as η_1 is known) but faces *distortion* (as η_1 affects his effort incentives). HM thus assume that the cost of effort is in financial terms so that it also declines with high η_1. We instead address distortion by the shape of the contract: It is convex, via the v^{-1} transformation. If noise is high, the contract gives a greater number of dollars for exerting effort ($\partial c / \partial r_1$), to exactly offset the lower marginal felicity of each dollar ($v'(c)$). Therefore, the marginal felicity from effort remains $v'(c)\partial c / \partial r_1 = g'\left(a_1^*\right)$, and incentives are preserved regardless of $\underline{u}$ or η_1. Allowing for convex contracts enables us to dispense with a pecuniary cost of effort. If the cost of effort is pecuniary ($v(c) = c$),

$v^{-1}(c) = c$ and so no transformation is needed. Since both the costs and benefits of effort are in monetary terms, a high η_1 reduces them equally. Thus, incentives are unchanged even with a linear contract.

We now move to $T > 1$. In all periods $t < T$, the agent is now exposed to risk, since he does not know future noise realizations when he chooses a_t. Much like the effect of a high current noise realization, if the agent expects future noise to be high, his incentives to exert effort are reduced. This would typically require the agent to integrate over future noise realizations when choosing a_t, leading to high complexity. Here, the unknown future noise outcomes do not matter, as can be seen in the heuristic proof. Before $T + 1$, η_{T+1} is unknown. However, (12) shows that the unknown η_{T+1} enters additively and does not affect the incentive constraints of the $t = 1, ..., T$ problems—regardless of what η_{T+1} turns out to be, the contract must give the agent a marginal felicity of $g'(a_t^*)$ for exerting effort at t.[11] Our timing assumption thus allows us to solve the multiperiod problem via backward induction, reducing it to a succession of one-period problems, each of which can be solved tractably. Equation (7) shows that, if the target action (and thus marginal cost of effort) is constant, incentives must be constant time-by-time as well as state-by-state, and so only aggregate performance ($r = \sum_{t=1}^{T} r_t$) matters.

Even though we can consider each problem separately, the periods remain interdependent. Much like the current noise realization, past outcomes may affect the current effort choice. The Mirrlees (1974) contract punishes the agent if final output is below a threshold. Therefore, if the agent can observe past outcomes, he will shirk if interim output is high. This complexity distinguishes our multiperiod model from a static multi-action model, where the agent chooses T actions simultaneously. Unlike in a multi-action model, here the agent observes past outcomes when taking his current action, and can vary his action in response. HM assume exponential utility and a pecuniary cost of effort to remove such "wealth effects" and eliminate the intertemporal link between periods. We instead ensure that past outcomes do not distort incentives via the above v^{-1} transformation, and so do not require either assumption.

Appendix B rules out randomized contracts. There are two effects of randomization. First, it leads to inefficient risk-sharing, for any concave u. Second, changing the reward for effort from a certain payment to a lottery may increase or decrease his effort incentives.[12] We show that with NIARA utility, this second effect is negative. Thus, both effects of randomization are

---

[11] This can be most clearly seen in the definition of the new utility function (13), which "absorbs" the $T + 1$ period problem.

[12] With separable utility, it is simple to show that randomization is inefficient, and so the principal offers the least risky contract that achieves incentive compatibility. With non-separable utility, introducing additional randomization via a riskier contract than necessary may be desirable (an example of the theory of second best); if low effort leads to a random payoff, this may induce the agent to increase effort. Gjesdal (1982) and Arnott and Stiglitz (1988) derive sufficient conditions under which randomization is suboptimal. Our conditions to guarantee the suboptimality of random contracts generalize their results to broader agency problems (their setting focuses on insurance).

undesirable, and deterministic contracts are unambiguously optimal. The proof makes use of the independence of noises and the log-concavity of $\eta_2, ..., \eta_T$. While these assumptions, combined with NIARA utility, are sufficient to rule out randomized contracts, they may not be necessary. In future research, it would be interesting to explore whether randomized contracts can be ruled out in broader settings.[13]

In addition to allowing for stochastic contracts, Theorem 1 also allows for $a_t^* = \overline{a}$, under which the IC constraint is an inequality. Therefore, the contract in (6) only provides a lower bound on the contract slope. A sharper-than-necessary contract has a similar effect to a stochastic contract, since it subjects the agent to additional risk. Again, the combination of NIARA and independent and log-concave noises is sufficient to rule out such contracts.

If the analysis is restricted to deterministic contracts and $a_t^* < \overline{a} \ \forall \ t$, the contract in (6) is the only incentive-compatible contract (for the signal values realized on the equilibrium path). We can thus relax the above three assumptions. This result is stated in Remark 1 below.

Remark 1. **(Optimal Deterministic Contract, $a_t^* < \overline{a} \ \forall \ t$)** Consider only deterministic contracts and $a_t^* < \overline{a} \ \forall \ t$. Relax the assumptions of NIARA utility, independent noises, and log-concave noises for $\eta_2, ..., \eta_T$. Any incentive-compatible contract takes the form

$$c = v^{-1}\left(\sum_{t=1}^{T} g'\left(a_t^*\right) r_t + K\right), \qquad (16)$$

where K is a constant. The optimal deterministic contract features a constant K that makes the agent's participation constraint bind.

Proof. See Appendix B. ∎

Remark 2 states that the contract's incentive compatibility is robust to the timing assumption. In particular, if noise follows the action in each period, the contract in Theorem 1 continues to implement the target actions—since it provides sufficient incentives state-by-state, it automatically does so on average. However, we can no longer show that it is optimal, since there are many other contracts that provide sufficient incentives on average.

Remark 2. **(Robustness of the Contract's Incentive Compatibility to Timing)** For any timing of the noise $(\eta_t)_{t=1...T}$ (i.e., regardless of whether it follows or precedes a_t in each period), the contract in Theorem 1 is incentive

---

[13] For instance, consider $T = 2$. We only require that $\widehat{u}(x)$ as defined in (46) exhibits NIARA. The concavity of η_2 is sufficient, but unnecessary for this. Separately, if NIARA is violated, effort incentives rise with randomization. However, this effect may be outweighed by the inefficient risk-sharing, so randomized contracts may still be dominated.

compatible and implements $(a_t^*)_{t=1,...,T}$. Indeed, given the contract, the agent's utility is

$$u\left(\sum_{t=1}^{T} g'\left(a_t^*\right)(a_t + \eta_t) + K - \sum_{t=1}^{T} g\left(a_t\right)\right),$$

so that, regardless of the timing of $(\eta_t)_{t=1...T}$, the agent maximizes his utility by taking action $a_t = a_t^*$, as it solves $\max_{a_t} g'\left(a_t^*\right) a_t - g\left(a_t\right)$.

Closed-form solutions allow the economic implications of a contract to be transparent. We close this section by considering two specific applications of Theorem 1 to executive compensation, to highlight the implications that can be gleaned from a tractable contract structure. The firm's log equity return is the natural choice of signal r for CEOs, since they are agents of shareholders. When the cost of effort is pecuniary ($v(c) = c$), Theorem 1 implies that the CEO's dollar pay c is linear in the firm's return r. Hence, the relevant incentives measure is the dollar change in CEO pay for a given percentage change in firm value (i.e., "dollar-percent" incentives), as advocated by Hall and Liebman (1998).

Another common specification is $v(c) = \ln c$, in which case the CEO's utility function (2) now becomes, up to a monotonic (logarithmic) transformation,

$$\mathrm{E}\left[U\left(ce^{-g(a)}\right)\right] \geq \underline{U}, \qquad (17)$$

where $u(x) \equiv U(e^x)$ and $\underline{U} \equiv \ln \underline{u}$ is the CEO's reservation utility. Utility is now multiplicative in effort and cash; Edmans, Gabaix, and Landier (2009) show that multiplicative preferences are necessary to generate empirically consistent predictions for the scaling of various measures of CEO incentives with firm size. Thus, the ability to drop the HM assumption of $v(c) = c$ becomes valuable. Applying Theorem 1 with $T = 1$ for simplicity, the optimal contract becomes

$$\ln c = g'(a^*)r + K. \qquad (18)$$

The contract prescribes the percentage change in CEO pay for a percentage change in firm value, i.e., "percent-percent" incentives; the level of incentives $g'(a^*)$ is independent of the utility function U and the noise distribution. Murphy (1999) advocated this elasticity measure over alternative incentive measures (such as "dollar-percent" incentives) on two empirical grounds: It is invariant to firm size, and firm returns have much greater explanatory power for percentage than dollar changes in pay. However, he notes that "elasticities have no corresponding agency-theoretic interpretation." The above analysis shows that elasticities are the theoretically justified measure under multiplicative preferences, for any utility function. This result extends the work of Edmans, Gabaix, and Landier (2009), who advocated "percent-percent" incentives in a risk-neutral, one-period model.

A second advantage of tractability is that it allows us to study whether executive contracts should be convex and thus contain stock options. Dittmann and Maug (2007) calibrate the standard CARA model and show that it predicts concave contracts;[14] hence, some commentators (e.g., Bebchuk and Fried 2004) argue that the use of options is evidence that CEO compensation is inefficient. Our closed-form solutions give the shape of the contract analytically, without the need for calibration. Options are convex in firm value, rather than firm returns. If S denotes firm value, we have $r = \ln S$. Thus, (18) becomes

$$c(S) = e^{K} S^{g'(a^*)}. \tag{19}$$

The convexity of the contract depends only on the marginal cost of effort, $g'(a^*)$. If it exceeds unity, the contract is convex in S. More broadly, while the HM setup delivers linear contracts, our setting can also accommodate convex and concave contracts.

1.2 Continuous time

This section shows that the contract has the same form in continuous time, where actions and noise are simultaneous. In the continuous-time version of the model, at every instant t, the agent takes action a_t and the principal observes signal r_t, where

$$r_t = \int_0^t a_s ds + \eta_t, \tag{20}$$

$\eta_t = \int_0^t \sigma_s dZ_s + \int_0^t \mu_s ds$, Z_t is a standard Brownian motion, and $\sigma_t > 0$ and μ_t are deterministic. The agent's utility function is

$$\mathrm{E}\left[u\left(v(c) - \int_0^T g(a_t)\, dt \right) \right]. \tag{21}$$

The principal observes the path of $(r_t)_{t \in [0,T]}$ and wishes to implement a deterministic action $(a_t^*)_{t \in [0,T]}$ at each instant. She solves Program 1 with utility function (21). The optimal contract is given by Proposition 1, and of the same tractable form as Theorem 1.

Proposition 1. (Optimal Contract, Continuous Time) *The following contract is optimal. The agent is paid*

$$c = v^{-1}\left(\int_0^T g'(a_t^*)\, dr_t + K \right), \tag{22}$$

where K is a constant that makes the participation constraint bind
$(\mathrm{E}\left[u\left(\begin{array}{c} \int_0^T g'(a_t^*)\, dr_t + K \\ - \int_0^T g(a_t^*)\, dt \end{array} \right) \right] = \underline{u}).$

---

[14] Dittmann, Maug, and Spalt (2010) show that a loss-averse utility function can deliver convex contracts. Dittmann and Yu (2010) show that convex contracts are optimal if the agent also affects firm risk.

In particular, if the target action is time-independent ($a_t^* = a^* \; \forall \; t$), the contract

$$c = v^{-1} \left(g' \left(a^* \right) r_T + K \right) \tag{23}$$

is optimal.

Proof. See Appendix B. ∎

To highlight the link with the discrete-time case, consider the model of Section 1.1 and define $r = \sum_{t=1}^{T} r_t = \sum_{t=1}^{T} a_t + \sum_{t=1}^{T} \eta_t$. Taking the continuous time limit of Theorem 1 gives Proposition 1. This consistency suggests that, if reality is continuous time, it is best approximated in discrete time by modeling noise before effort in each period. A brief intuition is that, in continuous time, the agent has significant freedom compared to the principal's information, which restricts the set of admissible contracts. In discrete time where noise follows the action, the principal has greater information since she can subtract profits at time $t - 1$ from profits at time t to learn the profits earned at time t (Hellwig and Schmidt 2002), which in turn allows her to achieve the first-best (Mueller 2000). Modeling the noise before the action gives the agent the same freedom he has in continuous time, and thus leads to similar contracts. Biais et al. (2007) similarly show convergence between discrete and continuous time in a cash flow diversion model where noise occurs before the action.

1.3 Discussion: What is necessary for tractable contracts?

The framework considered thus far shows that tractable contracts can be achieved without requiring exponential utility, a pecuniary cost of effort, continuous time, or Gaussian noise. However, it has still imposed a number of restrictions. We now discuss the features that are essential for our contract structure, inessential features that we have already relaxed in extensions, and additional assumptions that may be relaxable in future research.

1. *Timing of noise.* This assumption is essential to attaining simple contracts in discrete time as it restricts the principal's flexibility. Remark 2 states that, if a_t precedes η_t, contract (6) still implements $\left(a_t^* \right)_{t=1,..,T}$. However, we can no longer show that it is optimal.

2. *Fixed target action.* The analysis thus far has focused on the cheapest contract to implement a given path of target actions. In Section 2.1, we allow the target action to depend on the current-period noise, and in Section 2.2, we derive a sufficient condition for the optimal target action to be deterministic.

3. *Risk-neutral principal.* The full proof of Theorem 1 extends the model to the case of a risk-averse principal. If the principal wishes to minimize $E[w(c)]$ (where w is an increasing function) rather than $E[c]$, then

contract (6) is optimal if $u \left(v \left(w^{-1} \left(\cdot \right) \right) - \sum_t g \left(a_t^* \right) \right)$ is concave. This holds if, loosely speaking, the principal is not too risk averse.

4. *NIARA utility, independent and log-concave noise.* Remark 1 states that, if $a_t^* < \overline{a} \ \forall \ t$ and deterministic contracts are assumed, (6) is the only incentive-compatible contract. Therefore, these assumptions are not required. Allowing for $a_t^* = \overline{a}$ and stochastic contracts, these assumptions are sufficient but may not be necessary.

5. *Unidimensional noise and action.* Appendix C shows that our model is readily extendable to settings where the action a and the noise η are multidimensional. A close analog to our result obtains.

6. *Linear signal,* $r_t = a_t + \eta_t$. Remark 3 in Section 2.1 shows that with general signals $r_t = R \left(a_t, \eta_t \right)$, the optimal contract remains tractable and its functional form remains independent of u, $\underline{u}$, and the distribution of η.

7. *Timing of consumption.* The current setup assumes that the agent only consumes at the end of period T. Edmans et al. (2011) develop the analog of Theorem 1 where the agent consumes in each period, for the case of $v \left(c \right) = \ln c$ and a CRRA utility function.

8. *Renegotiation.* With a noise-independent action, there is no scope for renegotiation after the agent observes the noise. With a noise-dependent action, since the contract specifies an optimal action for every realization of η, again there is no incentive to renegotiate.

2. Optimal Contract with Noise-dependent Actions

Thus far, we have considered contracts that implement a fixed effort level, independent of the realized noise. Section 2.1 allows for the principal to implement an action that depends on the current period noise, similar to models in which the agent observes a state of nature before choosing his action. Section 2.2 derives a sufficient condition under which the optimal action is independent of the current period noise, and thus the focus on a fixed target action in Section 1 solves the full contracting problem.

2.1 Contingent target actions

Suppose the principal now wishes to implement the "action function" $A_t \left(\eta_t \right)$, which defines the target action for each noise realization. (Thus far, we have assumed $A_t \left(\eta_t \right) = a_t^*$.) Since different noises η_t may lead to the same observed signal $r_t = A_t \left(\eta_t \right) + \eta_t$, the analysis must consider revelation mechanisms and messages. If the agent announces noises $\widehat{\eta}_1, ..., \widehat{\eta}_T$, he is paid $c = C \left(\widehat{\eta}_1, ..., \widehat{\eta}_T \right)$ if the observed signals are $A_1 \left(\widehat{\eta}_1 \right) + \widehat{\eta}_1, ..., A_T \left(\widehat{\eta}_T \right) + \widehat{\eta}_T$, and a very low amount $\underline{c}$ otherwise.

As in the core model, we focus on $A_t \left(\eta_t \right) > \underline{a} \ \forall \ \eta_t$, else a flat contract would be optimal for some noise realizations. We make three additional technical

assumptions: The action space $\mathcal{A}$ is open, $A_t(\eta_t)$ is bounded within any compact subinterval of η, and $A_t(\eta_t)$ is almost everywhere continuous. The final assumption still allows for a countable number of jumps in $A_t(\eta_t)$. Given the complexity and length of the proof that randomized contracts are inferior in Theorem 1, we now restrict the analysis to deterministic contracts and assume $A_t(\eta_t) < \overline{a}$. We conjecture that the same arguments in that proof continue to apply with a noise-dependent target action.

The optimal contract induces both the target effort level ($a_t = A_t(\eta_t)$) and truth-telling ($\widehat{\eta}_t = \eta_t$). It is given by Proposition 2:

Proposition 2.   (Optimal Contract, Noise-dependent Action) A series of contingent actions $(A_t(\eta_t))_{t=1\dots T}$ can be implemented if and only if for all t, $A_t(\eta_t) + \eta_t$ is nondecreasing in η_t. If this condition is satisfied, the following contract is optimal. For each t, after noise η_t is realized, the agent communicates a value $\widehat{\eta}_t$ to the principal. If the subsequent signal is not $A_t(\widehat{\eta}_t) + \widehat{\eta}_t$ in each period, he is paid a very low amount $\underline{c}$. Otherwise, he is paid $C(\widehat{\eta}_1, ..., \widehat{\eta}_T)$, where

$$C(\eta_1, ..., \eta_T) = v^{-1}\left(\sum_{t=1}^{T} g(A_t(\eta_t)) + \sum_{t=1}^{T}\int_{\underline{\eta}}^{\eta_t} g'(A_t(x))\,dx + K\right),$$
(24)

$\underline{\eta}$ is an arbitrary constant, and K is a constant that makes the participation constraint bind ($\mathrm{E}\left[u\left(\sum_{t=1}^{T}\int_{\underline{\eta}}^{\eta_t} g'(A_t(x))\,dx + K\right)\right] = \underline{u}$).

Proof. (Heuristic.) Appendix B presents a rigorous proof that does not assume differentiability of V and A. Here, we give a heuristic proof that conveys the essence of the result using first-order conditions. We set $T = 1$ and drop the time subscript.

Instead of reporting η, the agent could report $\widehat{\eta} \neq \eta$, in which case he receives $\underline{c}$ unless $r = A(\widehat{\eta}) + \widehat{\eta}$. Therefore, he must take action a such that $\eta + a = \widehat{\eta} + A(\widehat{\eta})$, i.e., $a = A(\widehat{\eta}) + \widehat{\eta} - \eta$. In this case, his utility is $V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta)$. The truth-telling constraint is thus

$$\eta \in \arg\max_{\widehat{\eta}} V(\widehat{\eta}) - g(A(\widehat{\eta}) + \widehat{\eta} - \eta).$$

The first-order condition is

$$V'(\eta) = g'(A(\eta))\,A'(\eta) + g'(A(\eta)).$$

Integrating over η gives the indirect felicity function

$$V(\eta) = g(A(\eta)) + \int_{\underline{\eta}}^{\eta} g'(A(x))\,dx + K$$

for constants $\underline{\eta}$ and K. The associated pay is given by (24). ∎

The contract in Proposition 2 remains in closed form, and its functional form does not depend on u, $\underline{u}$ nor the distribution of η.[15] However, it is somewhat more complex than the contracts in Section 1, as it involves calculating an integral. Proposition 2 also identifies the class of action functions that is implementable. An action function is implementable if and only if $A_t(\eta_t) + \eta_t$ is nondecreasing in η_t. If this condition is not satisfied, and a higher noise corresponds to a significantly lower action, the agent would over-report the noise and exert less effort.

Remark 3 extends Proposition 2 to general signals.

Remark 3.   (Extension of Proposition 2 to General Signals) Suppose the signal is a general function $r_t = R(a_t, \eta_t)$, where R is differentiable and has positive derivatives in both arguments, $R_1(a, \eta)/R_2(a, \eta)$ is nondecreasing in a, and $R(A_t(\eta_t), \eta_t)$ is nondecreasing in η_t. The same analysis as in Proposition 2 derives the following contract as optimal:

$$
C(\eta_1, ..., \eta_T) = v^{-1}\left(\sum_{t=1}^{T} g(A_t(\eta)) \right.
$$
$$
\left. + \int_{\underline{\eta}}^{\eta_t} g'(A_t(x)) \frac{R_2(A_t(x), x)}{R_1(A_t(x), x)} dx + K \right), \qquad (25)
$$

where $\underline{\eta}$ is an arbitrary constant and K is a constant that makes the participation constraint bind.

The heuristic proof is as follows (setting $T = 1$ and dropping the time subscript). If η is observed and the agent reports $\widehat{\eta} \neq \eta$, he has to take action a such that $R(a, \eta) = R(A(\widehat{\eta}), \widehat{\eta})$. Taking the derivative at $\widehat{\eta} = \eta$ yields $R_1 \partial a/\partial\widehat{\eta} = R_1 A'(\eta) + R_2$. The agent solves $\max_{\widehat{\eta}} V(\widehat{\eta}) - g(a(\widehat{\eta}))$, with first-order condition $V'(\eta) - g'(A(\eta)) \partial a/\partial\widehat{\eta} = 0$. Substituting for $\partial a/\partial\widehat{\eta}$ from above and integrating over η yields (25).

2.2 Sufficient conditions for optimal effort to be deterministic

The analysis has thus far focused on the optimal implementation of a given path of actions or action functions. Solving the full contracting problem—the choice of the optimal action, in addition to its implementation—is typically highly complex. It can usually only be solved by restricting the action space to being binary, as this allows derivation of a simple condition to guarantee that the high effort level is optimal. This formulation is used by Holmstrom and Tirole (1998), He (2009, 2011), and Biais et al. (2010) and is the canonical model laid out in Tirole's (2005) textbook. With a continuous effort decision, the full contracting problem is usually intractable as there is a continuum of possible

---

[15] Even though (24) features an integral over the support of η, it does not involve the distribution of η.

effort choices. As a result, many contracting theories focus on implementing a given effort level, and empirical analyses typically calibrate to a given effort level. This section derives a class of settings in which the optimal action is indeed fixed, and thus the full contracting problem can be solved tractably under a continuous action space.

We consider the optimal action function $A(\eta)$, specializing to $T = 1$ for simplicity and dropping the time index. The principal chooses $A(\eta)$ to maximize

$$\max_{\{a(\eta)\}} \int b(a(\eta), \eta) f(\eta) d\eta - C[A]. \tag{26}$$

The first term represents the productivity of effort, where $a(\eta) = \min\left(A(\eta), \overline{\overline{a}}\right)$ and $\overline{\overline{a}} < \overline{a}$ is the highest productive effort level, representing full productive efficiency. The $\min\left(A(\eta), \overline{\overline{a}}\right)$ function conveys the fact that, while the action space may be unbounded ($\overline{a}$ may be infinite), there is a limit to the number of productive activities the agent can undertake to benefit the principal. In a cash flow diversion model, $\overline{\overline{a}}$ reflects zero stealing; in an effort model, there is a limit to the number of hours a day the agent can work while remaining productive; in a project selection model, there is a limit to the number of positive-NPV projects available: $\overline{\overline{a}}$ reflects taking all of these projects while rejecting negative-NPV projects. In addition to being economically realistic, this assumption is useful technically as it prevents the optimal action from being infinite. Actions $a > \overline{\overline{a}}$ do not benefit the principal, but improve the signal: One interpretation is manipulation (see Appendix F). Clearly, the principal will never wish to implement $a > \overline{\overline{a}}$. We will refer to $\overline{\overline{a}}$ as "high effort," to use similar terminology to models with discrete effort levels (e.g., high, medium, low) in which the high effort level is typically optimal. $b(\cdot)$ is the productivity function of effort, which is differentiable with respect to $a(\eta)$. $f(\eta)$ is the density of η, assumed to be finite. The second term, $C[A]$, is the expected cost of the contract required to implement $A(\eta)$ (we suppress the dependence on η for brevity).

We assume that g is strictly convex and that $g \circ (g')^{-1}$ and g' are convex.[16] Lemma 1 bounds the difference in the costs of the contract implementing high effort (denoted $C\left[\overline{\overline{A}}\right]$), and an arbitrary contract:

Lemma 1. **(Bound on Difference in Costs)** There exists a function $\lambda(\overline{\overline{a}}, \eta)$ such that, for all plans $\{a(\eta)\}$ where $\forall \eta, a(\eta) \leq \overline{\overline{a}}$,

$$C\left[\overline{\overline{A}}\right] - C[A] \leq \int \lambda(\overline{\overline{a}}, \eta)(\overline{\overline{a}} - a(\eta)) d\eta. \tag{27}$$

Proof. See Appendix B. ∎

---

[16] These assumptions are satisfied for many standard cost functions (e.g., $g(a) = Ga^2$ and $g(a) = e^{Ga}$ for $G > 0$).

Theorem 2 gives conditions under which high effort is optimal for all noise realizations: the "high effort principle".

Theorem 2. (High Effort Principle) Assume that $\forall \eta, \forall a \leq \overline{\overline{a}}, \partial_1 b(a, \eta)$ $f(\eta) \geq \lambda(\overline{a}, \eta)$, i.e., the marginal benefit of effort is sufficiently large. Then, the optimal plan is to implement high effort, $A(\eta) = \overline{\overline{a}}$.

Proof. For any plan,

$$\int \left(b(\overline{a}, \eta) - b(a(\eta), \eta) \right) f(\eta) \, d\eta \geq \int \inf_a \partial_1 b(a, \eta) \left(\overline{a} - a(\eta) \right) f(\eta) \, d\eta$$

$$\geq \int \lambda(\overline{a}, \eta) \left(\overline{a} - a(\eta) \right) d\eta$$

$$\geq C\left[\overline{\overline{A}}\right] - C[A]$$

by Lemma 1. Hence,

$$\int b(\overline{a}, \eta) f(\eta) \, d\eta - C\left[\overline{\overline{A}}\right] \geq \int b(a(\eta), \eta) f(\eta) \, d\eta - C[A],$$

i.e., the principal's objective is maximized by inducing high effort. ∎

Theorem 2 shows that, if the marginal benefit of effort is sufficiently greater than the marginal cost, then high effort is optimal. A sufficient (although unnecessary) condition is for the firm to be sufficiently large. To demonstrate this, we parameterize the b function by $b(a, \eta) = Sb_*(a, \eta)$, where S is the baseline value of the output under the agent's control. For example, if the agent is a CEO, S is firm size; if he is a divisional manager, S is the size of his division. We will refer to S as firm size for brevity. Under this specification, the benefit of effort is multiplicative in output. This is plausible for most agent actions, which can be "rolled out" across the whole company and thus have a greater effect in a larger firm, such as the choice of strategy or reorganizing production to cut costs.[17] Let $\overline{F}$ denote the complementary cumulative distribution function of η, i.e., $\overline{F}(x) = \Pr(\eta \geq x)$. We assume that $\sup_\eta \overline{F}(\eta)/f(\eta) < \infty$ and $\inf_\eta \partial_1 b_*(\overline{a}, \eta) > 0$, and define

$$S_* = \frac{\Lambda(\overline{\overline{a}})}{\inf_\eta \partial_1 b_*(\overline{\overline{a}}, \eta)}, \qquad \Lambda(\overline{\overline{a}}) \equiv \frac{g'(\overline{\overline{a}}) + g''(\overline{\overline{a}}) \sup_\eta \frac{\overline{F}(\eta)}{f(\eta)}}{v'\left(v^{-1}\left(u^{-1}(\underline{u}) + g(\overline{\overline{a}}) + (\overline{\eta} - \underline{\eta})g'(\overline{\overline{a}})\right)\right)}.$$

$$(28)$$

---

[17] Bennedsen, Perez-Gonzalez, and Wolfenzon (2009) provide empirical evidence that CEOs have the same percentage effect on firm value, regardless of firm size; Edmans, Gabaix, and Landier (2009) show that a multiplicative production function is necessary to generate empirically consistent predictions for the scaling of various measures of incentives with firm size.

Appendix D.2 shows that, if $S > S_*$, i.e., the firm is sufficiently large, then it is optimal for the principal to induce high effort. Indeed, in Theorem 2, we can take $\lambda\left(\overline{\overline{a}}, \eta\right) = \Lambda\left(\overline{\overline{a}}\right) f(\eta)$.

The intuition for the above is as follows. The numerator of $\Lambda\left(\overline{\overline{a}}\right)$ contains the two costs of inducing high effort—disutility (the first term) plus the risk imposed by the contract required to implement effort (the second term). These are scaled by the denominator, where the term in brackets is an upper bound on the agent's pay. The costs of effort are thus of similar order of magnitude to the agent's wage. The benefit of effort is enhanced firm value and thus of similar order of magnitude to firm size. If the firm is sufficiently large ($S > S_*$), the benefits of effort outweigh the costs and so high effort is optimal. For example, consider a firm with a \$10b market value and conservatively assume that high effort increases firm value by only 1%. Then, high effort creates \$100m of value, which vastly outweighs the agent's salary. Even if it is necessary to double the agent's salary to compensate him for the costs of increased effort, this is swamped by the benefits.

The comparative statics on the threshold firm size S_* are intuitive. First, S_* is increasing in noise dispersion, because the firm must be large enough for high effort to be optimal for all noise realizations. Indeed, a rise in $\overline{\eta} - \underline{\eta}$ increases $u^{-1}(\underline{u}) + g(\overline{\overline{a}}) + (\overline{\eta} - \underline{\eta})g'(\overline{\overline{a}})$, lowers α, and raises $\sup \overline{F}/f$.[18] Second, it is increasing in the agent's risk aversion, parameterized by v. Third, it is increasing in the disutility of effort, and thus the marginal cost of effort $g'\left(\overline{\overline{a}}\right)$ and the convexity $g''(\overline{\overline{a}})$. Fourth, it is decreasing in the marginal benefit of effort ($\inf_\eta \partial_1 b_*\left(\overline{\overline{a}}, \eta\right)$).

We conjecture that the high effort principle applies in more general settings than those considered above. For instance, it likely continues to hold if $\overline{a}$ (the highest feasible effort level) equals $\overline{\overline{a}}$ (the highest productive effort level). This slight variant is economically very similar, since the principal never wishes to implement $A(\eta) > \overline{\overline{a}}$, but is substantially more complicated mathematically, because the agent's action space now has boundaries and so the incentive constraints become inequalities. We leave this extension to future research. Hellwig (2007) shows that this reason alone is sufficient for a boundary effort level to be optimal, even without the condition on the benefit of effort featured here. Since the incentive constraints become inequalities, the principal has greater freedom in choosing the contract, which allows her to select a cheaper contract. Thus, high effort is optimal in settings even without a large benefit of effort. Edmans et al. (2011) extend the optimality of high effort to general T, for the case where $v(c) = \ln c$ (multiplicative preferences) and u is CRRA.

A number of prior papers assume a fixed effort level as it removes the need to solve for the optimal action and leads to substantial tractability. Theorem 2

---

[18] For example, if the noise is uniform, then $\sup \overline{F}/f = \overline{\eta} - \underline{\eta}$.

provides the conditions under which this focus is valid. For example, in their calibration of CEO contracts, Dittmann and Maug (2007) and Dittmann, Maug, and Spalt (2010) assume a given effort level. Since they study CEOs who have a large benefit of effort, Theorem 2 rationalizes this approach.

Appendix D considers other sufficient conditions required for Theorem 2 to hold, which do not assume the benefit of effort is multiplicative in firm size. In addition, it shows that even if Theorem 2 does not hold, the optimal $\{A(\eta)\}$ can still be derived if we have a linear cost function; in this case, the optimal action is interior.

2.3 Determinants of the implemented effort level

The previous section assumed that productive capacity $\overline{\overline{a}}$ is exogenous. This section allows the principal to choose it endogenously according to the environment. We extend the contracting game to two stages. In the first stage, the principal chooses $\overline{\overline{a}}$. In practice, this may be achieved by physical investment or training the agent, for example. Since these actions are costly to reverse, we model them as irreversible. In the second stage, the game studied in Section 2.1 is played out. The action a may respond to the noise η, but the highest productive effort $\overline{\overline{a}}$ has been fixed.

The principal's payoff is

$$\int b\left(\min\left(A\left(\eta\right),\overline{\overline{a}}\right),\eta,\overline{\overline{a}}\right)d\eta - C\left[A\right], \tag{29}$$

where $b\left(a,\eta,\overline{\overline{a}}\right)$ is weakly increasing in a and decreasing in $\overline{\overline{a}}$. Higher productive capacity $\overline{\overline{a}}$ is costly to the principal. Maximizing (29) appears complex since the principal must choose both capacity $\overline{\overline{a}}$ and the action function $A\left(\eta\right)$. This section shows that, under certain conditions, the principal's problem can be simplified to one in which she chooses only capacity $\overline{\overline{a}}$.

We consider the two following problems.

Problem 1: Maximize over $\overline{\overline{a}}$ and all unrestricted contracts:

$$\max_{\overline{\overline{a}},\{A(\eta)\}} E\left[b\left(\min\left(A\left(\eta\right),\overline{\overline{a}}\right),\eta,\overline{\overline{a}}\right)\right] - C\left[A\right].$$

Problem 2: Maximize over $\overline{\overline{a}}$ and use the contract in Theorem 1 that implements $\overline{\overline{a}}$:

$$\max_{\overline{\overline{a}}} B\left(\overline{\overline{a}}\right) - C\left[\overline{\overline{a}}\right],$$

where $B\left(a\right) = E\left[b\left(a,\eta,a\right)\right]$ is the principal's expected payoff given target effort a, and $C\left[a\right]$ is the expected cost of the contract implementing a constant action a.

Problem 2 optimizes over only a scalar $\overline{\overline{a}}$, while Problem 1 optimizes over a whole continuum of contracts, including those that do not implement high

effort. However, under some simple conditions, both problems have the same solution—the principal cannot improve on implementing high effort. This result is shown in Theorem 3.

Theorem 3.   (Optimal Target Effort via a Two-stage Game) Let a^{**} denote the value of $\bar{\bar{a}}$ in a solution to Problem 1, and assume that $a^{**} > \underline{a}$ and that $\forall \eta$, $\inf_a \partial_1 b (a, \eta, a^{**}) f (\eta) \geq \lambda (a^{**}, \eta)$. Then, the solution of Problem 1 is the same solution as Problem 2: The solution of the problem that implements $A (\eta) = \bar{\bar{a}}$ is also the solution of the unrestricted contract. Thus, the principal's problem can be reduced to solving the optimization

$$\max_a B (a) - C [a].\tag{30}$$

Problem (30) optimally solves for a single deterministic target effort level a, rather than a state-dependent target action $A (\eta)$.

Proof. Immediate given Theorem 2. At a^{**}, the principal wishes to implement high effort, i.e., $a (\eta) = a^{**}$ for all η.   ■

The meaning of Theorem 3 is as follows. In the second stage, the principal wishes to implement the contract in Theorem 1 with $a^{*} = \bar{\bar{a}}$. In the first stage, when choosing $\bar{\bar{a}}$, she trades off the costs and benefits of higher $\bar{\bar{a}}$. For instance, in the examples at the end of this section, we have the standard result that the implemented effort level $\bar{\bar{a}}$ is decreasing in the agent's disutility and risk aversion, along with the noise dispersion. Since the principal knows the effort level $\bar{\bar{a}}$ will be chosen in the second stage, a potentially complex problem where she has to optimize over productive capacity $\bar{\bar{a}}$ and the action function $A (\eta)$ reduces to a simple problem where she chooses $\bar{\bar{a}}$ alone, i.e., (30).

A tradeoff exists in the first stage because the costs and benefits of flexibility are of similar order of magnitude. For example, increasing plant size has a continuous effect on firm value and involves a significant cost, which is also a function of firm size. However, a tradeoff does not exist in the second stage because the costs of effort are now a function of the agent's salary, and the benefits are discontinuous. Once the plant has been built, the agent must run it with full efficiency to prevent significant value loss—even small imperfections will cause large reductions in value and so the marginal benefit of effort is high (analogous to Kremer's 1993 O-ring theory). Thus, this enriched game features a simple optimal contract (since the target action in the second stage is constant), but one that also responds to the comparative statics of the environment. It may therefore be a potentially useful way of modeling various economic problems, to achieve tractability while at the same time generating comparative statics.

To calculate $C [a]$, Theorem 1 gives the optimal contract as

$$c = v^{-1} \left(g' (a) r + K \right),$$

where K satisfies $E\left[u\left(g'(a)r - g(a) + K\right)\right] = \underline{u}$. The expected cost of the contract is

$$C[a] \equiv E[c(r)] = E\left[v^{-1}\left(g'(a)r + K\right)\right].$$

It is straightforward to show that $C[a]$ increases in target effort a, reservation utility $\underline{u}$, and the dispersion of noise η; the proof relies on the dispersion techniques illustrated in Appendix A. The objective function (30) now becomes

$$\max_a B(a) - E\left[v^{-1}\left(g'(a)r + K\right)\right]. \tag{31}$$

This is a simple problem in many applied settings. We consider two examples below, deriving the contract explicitly and studying the comparative statics. In addition, in Appendix D, we provide a specialization of the conditions in Theorem 3 to these two examples. This allows straightforward verification that the optimal policy is indeed deterministic. At first glance, the condition in Theorem 3 may appear complex, since verifying it requires solving Problem 1. However, sufficient conditions are simply $\inf_a \partial_1 b(a, \eta, a^{**}) f(\eta) \geq \lambda(a^{**}, \eta)$ for all a^{**} and η. The value λ can be calculated up to an integral, so bounds are reasonably straightforward to check in a given setting, such as the following examples. Appendix D also analyzes a third example with CARA utility and a pecuniary cost of effort.

Example 1. CRRA and multiplicative preferences. Consider $v(x) = \ln x$ and $u(x) = e^{(1-\gamma)x}/(1-\gamma)$ for $\gamma > 1$, and redefine the cost function as $Gg(a)$, where G is a scalar that parameterizes the cost of effort. The agent's utility is now

$$E\left[\frac{\left(ce^{-Gg(a)}\right)^{1-\gamma}}{1-\gamma}\right].$$

This utility function is commonly used in macroeconomics; it is CRRA in consumption and multiplicative in consumption and effort. We also assume $\eta \sim N\left(0, \sigma^2\right)$. Then, the contract is

$$c(r) = \exp\left(Gg'(a)r + K\right), \tag{32}$$

with

$$K = \ln \underline{c} + Gg(a) + (\gamma - 1)G^2 g'(a)^2 \sigma^2/2 - Gg'(a)a,$$

where $u\left(\ln \underline{c}\right)$ is the reservation utility.

We now study the optimal effort level. The expected cost of the contract is

$$C[a] = \underline{c}\exp\left(Gg(a) + \gamma G^2 g'(a)^2 \sigma^2/2\right).$$

Consider an additive benefit of effort, i.e., $b\left(a, \eta, \overline{\overline{a}}\right) = a + \left(a - \overline{\overline{a}}\right) \beta\left(\overline{\overline{a}}, \eta\right)$, for some function $\beta\left(\overline{\overline{a}}, \eta\right) \geq \lambda\left(\overline{\overline{a}}, \eta\right) / f(\eta)$. In the second stage of the game, the principal wishes to implement $\overline{\overline{a}}$ for all η, because the marginal cost of shirking (parameterized by β) is sufficiently high. Moving to the first stage, since the principal knows that $a = \overline{\overline{a}}$ in the second stage, her benefit function is $b\left(a, \eta, \overline{\overline{a}}\right) = a$; effort has an additive effect.

The principal maximizes

$$\max_a a - C\left[a\right]$$

with first-order condition

$$1 - \underline{c}e^{Gg(a) + \gamma\, G^2 g'(a)^2 \sigma^2 / 2}\left(Gg'\left(a\right) + \gamma\, G^2 g'\left(a\right) g''\left(a\right)\sigma^2\right) = 0.$$

Taking cross-partials shows that a is decreasing in the cost of effort G, risk σ, and risk aversion γ.[19] These are the same comparative statics as in HM. Moreover, since we do not require exponential utility nor a pecuniary cost of effort, we obtain additional implications. First, the contract is log-linear in returns, i.e., the relevant incentive measure is the percentage change in pay for the percentage change in firm value. Murphy (1999) argues that log-linear contracts are empirically more relevant. Second, the model can accommodate convex and concave contracts, rather than only linear ones. If $Gg'\left(a\right) > 1$, the contract is convex, as documented empirically for CEOs. Third, the implemented effort level a is decreasing in the agent's reservation utility $\underline{c}$ and thus his wealth. In HM, there are no comparative statics with respect to wealth, since the framework requires zero wealth effects. Our framework allows for wealth effects: With general concave $\upsilon\left(\cdot\right)$, richer agents are less motivated by money. Thus, they need to be given stronger incentives to prevent them from shirking; this increase in risk causes them to demand a greater risk premium. This leads to an interesting disadvantage of hiring wealthy agents—since they are already rich, they are more costly to incentivize. Thus, the principal optimally chooses a lower effort level. Edmans and Gabaix (2011) shows that this leads to distortions in CEO assignment and aggregate production in a market equilibrium.

Example 2. Additively separable utility. We now consider $u\left(x\right) = x$, so that utility is additively separable rather than multiplicative. We also consider $\upsilon\left(x\right) = x^\gamma$, where $\gamma \in (0, 1]$. We have $k = g\left(a\right) + \underline{u} + g'\left(a\right)a$, and the contract is

$$c\left(r\right) = \left(g'\left(a\right)r + g\left(a\right) + \underline{u} + g'\left(a\right)a\right)^{1/\gamma}.$$

---

[19] As is well known, with θ a parameter, since $\frac{\partial^2 C(a,\theta)}{\partial a \partial \theta} + \frac{\partial^2 C(a,\theta)}{\partial a^2}\frac{\partial a}{\partial \theta} = 0$ and $\frac{\partial^2 C(a,\theta)}{\partial a^2} > 0$, $\frac{\partial a}{\partial \theta}$ has the opposite sign to $\frac{\partial^2 C(a,\theta)}{\partial a \partial \theta}$.

The expected cost is

$$C[a] = E\left[\left(g'(a)\eta + g(a) + \underline{u}\right)^{1/\gamma}\right].$$

We consider an additive benefit of effort, $b\left(a, \eta, \overline{\overline{a}}\right) = a + \left(a - \overline{\overline{a}}\right)\beta\left(\overline{\overline{a}}, \eta\right)$, as in Example 1. The contract is linear if $\gamma = 1$; otherwise, it is convex. Simple calculations show that the target action is decreasing in the cost of effort G, risk σ, and risk aversion γ.[20]

3. Conclusion

This article has developed a framework in which the optimal contract is tractable, without requiring exponential utility, a pecuniary cost of effort, or Gaussian noise. Two conditions are sufficient for tractability. The first is the focus on a deterministic target action, which is optimal if the agent has a large effect on firm outcomes (such as a CEO). The second is modeling either in continuous time or specifying the noise before the agent's action in discrete time. When these conditions hold, the contract's functional form is independent of the agent's utility function, reservation utility, and noise distribution. Furthermore, when the cost of effort can be expressed in financial terms, the contract is linear and so the slope, in addition to the functional form, is independent of these parameters.

Our article suggests several avenues for future research. The HM framework has proven valuable in many areas of applied contract theory owing to its tractability; however, some models have used the HM result in settings where the assumptions are not satisfied (see the critique of Hemmer 2004). Our framework allows simple contracts to be achieved in such situations. While we considered the specific application of executive compensation, other possible actions include bank regulation, team production, insurance, or taxation.

In addition, while our model has relaxed a number of assumptions required for tractability, it continues to impose a number of restrictions. For example, while Section 2.1 allows for the action to depend on the noise in period t, a useful extension would be to allow the action to depend on the full history of outcomes. This is a highly complex question related to the Mirrlees taxation problem, which is known to be difficult (see Farhi and Werning 2010 and Golosov, Troshkin, and Tsyvinski 2010 for recent analytical progress on those problems). Other restrictions are mostly technical rather than economic. For example, our multiperiod model assumes independent noises with log-concave density functions, and our extension to noise-dependent target actions assumes an open action set, where the highest feasible effort level exceeds the highest productive effort level. Further research may be able to broaden the current setup.

---

[20] A variant is the case $u(x) = x$ and $v(x) = \ln x$. Then, the contract is $\ln c(r) = g'(a)(r - a) + g(a) + \underline{u}$, and the expected cost is $C(a) = \exp\left[g(a) + \underline{u}\right] E\left[e^{g'(a)\eta}\right]$.

Appendix

All appendices are available online at http://www.sfsrfs.org.

References

Arnott, R., and J. Stiglitz. 1988. Randomization with Asymmetric Information. *RAND Journal of Economics* 19:344–62.

Baker, G. 1992. Incentive Contracts and Performance Measurement. *Journal of Political Economy* 100:598–614.

Baranchuk, N., G. Macdonald, and J. Yang. 2011. The Economics of Super Managers. *Review of Financial Studies*, forthcoming.

Bebchuk, L., and J. Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Compensation.* Cambridge, MA: Harvard University Press.

Bennedsen, M., F. Perez-Gonzalez, and D. Wolfenzon. 2009. Do CEOs Matter? Working Paper, Copenhagen Business School.

Biais, B., T. Mariotti, G. Plantin, and J. C. Rochet. 2007. Dynamic Security Design: Convergence to Continuous Time and Asset-pricing Implications. *Review of Economic Studies* 74:345–90.

Biais, B., T. Mariotti, J. C. Rochet, and S. Villeneuve. 2010. Large Risks, Limited Liability and Dynamic Moral Hazard. *Econometrica* 78:73–118.

Caplin, A., and B. Nalebuff. 1991. Aggregation and Social Choice: A Mean Voter Theorem. *Econometrica* 59:1–23.

Chassang, S. 2011. Calibrated Incentive Contracts. Working Paper, Princeton University.

Cooley, T., and E. Prescott. 2005. Economic Growth and Business Cycles. In Thomas Cooley (ed.), *Frontiers in Business Cycle Research.* Princeton, NJ: Princeton University Press.

DeMarzo, P., and M. Fishman. 2007. Optimal Long-term Financial Contracting. *Review of Financial Studies* 20:2079–127.

DeMarzo, P., and Y. Sannikov. 2006. Optimal Security Design and Dynamic Capital Structure in a Continuous-time Agency Model. *Journal of Finance* 61:2681–724.

Dittmann, I., and E. Maug. 2007. Lower Salaries and No Options? On the Optimal Structure of Executive Pay. *Journal of Finance* 62:303–43.

Dittmann, I., E. Maug, and O. Spalt. 2010. Sticks or Carrots? Optimal CEO Compensation When Managers Are Loss-averse. *Journal of Finance* 65:2015–50.

Dittmann, I., and K. C. Yu. 2010. How Important Are Risk-taking Incentives for Executive Compensation? Working Paper, Erasmus University–Rotterdam.

Edmans, A., and X. Gabaix. 2011. The Effect of Risk on the CEO Market. *Review of Financial Studies*, forthcoming.

Edmans, A., X. Gabaix, and A. Landier. 2009. A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium. *Review of Financial Studies* 22:4881–917.

Edmans, A., X. Gabaix, T. Sadzik, and Y. Sannikov. 2011. Dynamic CEO Compensation. Working Paper, University of Pennsylvania.

Farhi, E., and I. Werning. 2009. Capital Taxation: Quantitative Explorations of the Inverse Euler Equation. Working Paper, Harvard University.

———. 2010. Insurance and Taxation over the Life Cycle. Working Paper, Harvard University.

Garrett, D., and A. Pavan. 2010. Dynamic Managerial Compensation: On the Optimality of Seniority-based Schemes. Working Paper, Northwestern University.

Gjesdal, F. 1982. Information and Incentives: The Agency Information Problem. *Review of Economic Studies* 49:373–90.

Golosov, M., N. Kocherlakota, and A. Tsyvinski. 2003. Optimal Indirect Capital Taxation. *Review of Economic Studies* 70:569–87.

Golosov, M., M. Troshkin, and A. Tsyvinski. 2010. Optimal Dynamic Taxes. Working Paper, Yale University.

Grossman, S., and O. Hart. 1983. An Analysis of the Principal-agent Problem. *Econometrica* 51:7–45.

Hall, B., and J. Liebman. 1998. Are CEOs Really Paid Like Bureaucrats? *Quarterly Journal of Economics* 113:653–91.

Hall, B., and K. Murphy. 2002. Stock Options for Undiversified Executives. *Journal of Accounting and Economics* 33:3–42.

Harris, M., and A. Raviv. 1979. Optimal Incentive Contracts with Imperfect Information. *Journal of Economic Theory* 20:231–59.

He, Z. 2009. Optimal Executive Compensation When Firm Size Follows Geometric Brownian Motion. *Review of Financial Studies* 22:859–92.

———. 2011. Dynamic Compensation Contracts with Private Savings. *Review of Financial Studies*, forthcoming.

Hellwig, M. 2007. The Role of Boundary Solutions in Principal-agent Problems of the Holmstrom-Milgrom Type. *Journal of Economic Theory* 136:446–75.

Hellwig, M., and K. Schmidt. 2002. Discrete-time Approximations of the Holmstrom-Milgrom Brownian-motion Model of Intertemporal Incentive Provision. *Econometrica* 70:2225–64.

Hemmer, T. 2004. Lessons Lost in Linearity: A Critical Assessment of the General Usefulness of LEN Models in Compensation Research. *Journal of Management Accounting Research* 16:149–62.

Holmstrom, B., and P. Milgrom. 1987. Aggregation and Linearity in the Provision of Intertemporal Incentives. *Econometrica* 55:308–28.

Holmstrom, B., and J. Tirole. 1998. Private and Public Supply of Liquidity. *Journal of Political Economy* 106:1–40.

Laffont, J. J., and J. Tirole. 1986. Using Cost Observation to Regulate Firms. *Journal of Political Economy* 94:614–41.

Karatzas, I., and S. E. Shreve. 1991. *Brownian Motion and Stochastic Calculus*, 2nd edition. Berlin, Germany: Springer-Verlag.

Kremer, M. 1993. The O-ring Theory of Economic Development. *Quarterly Journal of Economics* 108:551–76.

Krishna, V., and E. Maenner. 2001. Convex Potentials with an Application to Mechanism Design. *Econometrica* 69:1113–19.

Lacker, J., and J. Weinberg. 1989. Optimal Contracts under Costly State Falsification. *Journal of Political Economy* 97:1345–63.

Landsberger, M., and I. Meilijson. 1994. The Generating Process and an Extension of Jewitt's Location Independent Risk Concept. *Management Science* 40:662–69.

Mirrlees, J. 1974. Notes on Welfare Economics, Information, and Uncertainty. In Michael Balch, Daniel McFadden, and Shih-Yen Wu (eds.), *Essays on Economic Behavior under Uncertainty*. Amsterdam, Netherlands: North-Holland Publishing.

Mueller, H. 2000. Asymptotic Efficiency in Dynamic Principal-agent Problems. *Journal of Economic Theory* 91:292–301.

Murphy, K. 1999. Executive Compensation. In Orley Ashenfelter and David Card (eds.), *Handbook of Labor Economics*, Vol. 3b. New York and Oxford: Elsevier/North-Holland Publishing.

Ou-Yang, H. 2003. Optimal Contracts in a Continuous-time Delegated Portfolio Management Problem. *Review of Financial Studies* 16:173–208.

Phelan, C., and R. Townsend. 1991. Private Information and Aggregate Behavior: Computing Multiperiod, Information-constrained Optima. *Review of Economic Studies* 58:853–81.

Prendergast, C. 2002. The Tenuous Tradeoff Between Risk and Incentives. *Journal of Political Economy* 110: 1071–102.

Rogerson, W. 1985. The First-order Approach to Principal-agent Problems. *Econometrica* 53:1357–68.

Rudin, W. 1987. *Real and Complex Analysis*, 3rd edition. New York: McGraw-Hill.

Sannikov, Y. 2008. A Continuous-time Version of the Principal-agent Problem. *Review of Economic Studies* 75:957–84.

Sappington, D. 1983. Limited Liability Contracts Between Principal and Agent. *Journal of Economic Theory* 29:1–21.

Shaked, M., and G. Shanthikumar. 2007. *Stochastic Orders*. Berlin, Germany: Springer-Verlag.

Shimer, R., and I. Werning. 2008. Liquidity and Insurance for the Unemployed. *American Economic Review* 98:1922–42.

Spear, S., and S. Srivastava. 1987. On Repeated Moral Hazard with Discounting. *Review of Economic Studies* 54:599–617.

Sung, J. 1995. Linearity with Project Selection and Controllable Diffusion Rate in Continuous-time Principal-agent Problems. *RAND Journal of Economics* 26:720–43.

Tirole, J. 2005. *The Theory of Corporate Finance*. Princeton, NJ: Princeton University Press.