The Rodney L. White Center for Financial Research

Dynamic Incentive Accounts

Alex Edmans
Xavier Gabaix
Tomasz Sadzik
Yuliy Sannikov

10-09

The Wharton School
University of Pennsylvania

Dynamic Incentive Accounts*

Alex Edmans — Wharton
Xavier Gabaix — NYU and NBER
Tomasz Sadzik — NYU
Yuliy Sannikov — Princeton

May 13, 2009

Abstract

Contracts in a dynamic model must address a number of issues absent from static frameworks. Shocks to firm value may weaken the incentive effects of securities given to the CEO (e.g. cause options to fall out of the money), and the impact of some CEO actions may not be felt until far in the future. To address these concerns, we derive the optimal contract in a setting where the CEO can affect firm value through both productive effort or costly manipulation, and may undo the contract by privately saving. The efficient contract takes a surprisingly simple form, and can be implemented by a "Dynamic Incentive Account." The CEO's expected pay is escrowed into an account, a fraction of which is invested in the firm's stock and the remainder in cash. The account features state-dependent rebalancing and time-dependent vesting. It is constantly rebalanced so that the equity fraction remains above a certain threshold; this threshold sensitivity is typically increasing over time even in the absence of career concerns. The account vests gradually both during the CEO's employment and after he quits, to deter short-termist actions before retirement.

Keywords: Contract theory, executive compensation, incentives, principal-agent problem, manipulation, private saving, vesting.

JEL Classification: D2, D3, G34, J3

*aedmans@wharton.upenn.edu, xgabaix@stern.nyu.edu, tsadzik@nyu.edu, sannikov@gmail.com.

For helpful comments, we thank Zhiguo He and Eric Talley, and seminar participants at the Harvard Law School / Sloan Foundation Conference on Corporate Governance, the NBER, NYU and the University of Chicago. Qi Liu and Andrei Savotchkine provided excellent research assistance.

1 Introduction

Many classical models of CEO compensation consider only a single period, or multiple unlinked periods. However, the optimal contract in a static analysis may be suboptimal in a dynamic world where the CEO's current actions, such as his effort or savings/consumption choice, impact future periods. For example, short-term contracts can encourage the CEO to manipulate earnings or scrap investment projects to boost the current stock price at the expense of fundamental value; these long-run costs may not appear until after the CEO has retired. By privately saving, the CEO can separate his consumption stream from the path of income provided by his contract, and thus undo the intended incentives. Securities given to incentivize the CEO may lose their power over time: if firm value declines, options fall out-of-the-money and bear little sensitivity to the stock price. In addition to the three above challenges, a dynamic setting also provides opportunities absent from a static framework – in particular, the firm has the option to reward current effort with future rather than contemporaneous compensation.

This paper analyzes optimal executive compensation in a dynamic model that allows for all of the above complexities, which are likely important features in real life. Despite the complications that result from a dynamic setting, the optimal contract is surprisingly simple and features intuitive economic principles. We first consider an infinite horizon model where the CEO has no option to manipulate earnings or privately save, to provide a benchmark against which to analyze the effect of introducing these complexities. In this baseline model, the optimal contract is time-independent: the sensitivity of pay to the firm's return is the same in each period. The relevant measure of incentives is the percentage change in CEO pay for a percentage change in firm value; translated into real variables, this is the fraction of CEO pay that comprises of stock. If the CEO's outside option doubles, his total pay doubles but the relative weighting on cash and stock remains the same. Thus, the contract is also scale-independent. This result extends to a dynamic setting Edmans, Gabaix and Landier (2009), who advocated this incentive measure in a one-period model with a risk-neutral CEO. The optimal contract also involves consumption smoothing. Since the agent is risk-averse, it is efficient to spread the reward for effort across all future periods rather than concentrating it in the current period (the "deferred reward principle"). This result is consistent with Boschen and Smith (1995), who find that firm performance has a much greater effect on the present value of future pay rather than on contemporaneous pay.

With a finite horizon, the sensitivity of income to firm returns is now increasing over time – the "increasing incentives principle." As the CEO approaches retirement, there are fewer periods in which to spread the reward for effort, and so the reward in the current period must increase. We thus generate a similar prediction to Gibbons and Murphy (1992), but without invoking career concerns.

Allowing the CEO to manipulate the stock price has two effects on the optimal contract, which must change to prevent such behavior. The CEO's income is now sensitive to firm returns even after retirement, to deter him from inflating the stock price just before he leaves. In addition, the contract sensitivity now must rise over time, even in an infinite-horizon model. This is because the CEO benefits immediately from short-termism as it boosts his current consumption, but the cost is only suffered in the future and thus has a discounted effect on the CEO's utility. Therefore, an increasing slope is needed to ensure that the CEO loses more dollars in the future than he gains today.

By contrast, the possibility of private savings does not change the contract's sensitivity to firm value, since it does not affect the CEO's effort. Instead, the ability to save privately affects the level of pay, causing it to increase more rapidly over time. Rising pay effectively saves for the CEO, thus removing the incentive for him to do so privately.

In practice, the optimal contract can be implemented in a straightforward manner. When initially appointed, the CEO is given a "Dynamic Incentive Account": a portfolio of which a given fraction is invested in the firm's stock and the remainder in cash. As time evolves, and firm value changes, this portfolio is constantly rebalanced, so that the fraction in the firm's stock remains sufficient to induce effort at minimum risk to the CEO. For example, a fall in the share price decreases the equity in the incentive account below the threshold fraction; this is addressed by using cash in the account to purchase stock. By contrast, if the stock appreciates, some of the equity can be sold without falling below the threshold, to reduce the risk borne by the CEO. The required fraction represents the contract's sensitivity, and so is constant in an infinite horizon model where manipulation is impossible, and increasing over time otherwise.

In addition to continuous rebalancing, the Dynamic Incentive Account also features gradual vesting, both during the CEO's employment and after his retirement. He can only withdraw a fraction of the account in each period, and it does not immediately vest upon leaving the firm – full withdrawal is only possible after a sufficient period has elapsed for the effects of manipulation to have been reversed. If the model horizon is infinite, the vesting fraction is time-independent (constant across periods), just like the contract sensitivity.

In sum, the Dynamic Incentive Account has two key features, which each achieve separate objectives. State-dependent rebalancing ensures that the CEO always exerts the required level of effort, while minimizing the risk that he bears. Time-dependent vesting ensures that the CEO always abstains from manipulation, while allowing him to finance consumption. The model thus offers theoretical guidance on how executive compensation might be reformed to address and prevent the problems that manifested in the recent crisis, at minimum cost. A number of commentators (e.g. Bebchuk and Fried (2004), Holmstrom (2005)) have argued that lengthening vesting horizons on stock and options may deter manipulation. Even if such a change could be achieved at little cost, it only solves one of the two problems: while it entails time-dependent vesting and thus addresses myopia, it does not involve state-dependent rebalancing and so ensure continued incentive compatibility over time.

Moreover, existing theories demonstrate costs of lengthening vesting horizons, which may lead to the optimal vesting horizon being short. Such costs arise because vesting and rebalancing are the same event in these models – therefore, long vesting prevents timely rebalancing and may aggravate the effort problem. In the Dynamic Incentive Account, vesting and rebalancing are separate events, allowing each issue to be addressed without worsening the other. For example, in Peng and Roell (2009), long vesting periods increase the risk borne by the CEO as they delay the rebalancing of stock for cash, and so the firm chooses a short vesting horizon even though this induces some manipulation. In our model, distant vesting can be achieved without imposing excessive risk: stock can be sold for cash upon good interim performance, although the proceeds are retained within the account. Brisley (2006) and Bhattacharyya and Cohn (2008) show that allowing the CEO to rebalance his securities for cash can increase his willingness to undertake risky projects by reducing his firm-specific risk. Since rebalancing can only be achieved through vesting, Bhattacharyya and Cohn show that the optimal vesting period is short. While they consider stock, Brisley analyzes options where rebalancing is only necessary upon strong performance, since only in-the-money options subject the CEO to risk. Therefore, as in our model, state-dependent rebalancing is optimal; since rebalancing and vesting are the same event in Brisley's model, this requires state-dependent vesting. Indeed, recent empirical studies (e.g. Bettis, Bizjak, Coles and Kalpathy (2008)) document that performance-based (i.e. state-dependent) vesting is becoming increasingly popular. However, state-dependent vesting may allow the CEO to manipulate the stock price upwards (an action not featured in the two above theories) and cash out his shares. Thus, state-dependent vesting has critically different effects to the combination of state-dependent rebalancing and time-dependent vesting – under our contract, high stock returns allow sales of equity, but the proceeds remain within the account in case the returns are subsequently reversed. Our framework incorporates manipulation and so requires these two features to achieve the two separate goals of effort inducement and manipulation deterrence.

In addition to the above papers on vesting horizons, our paper is also related to the literature on optimal contracts in the presence of manipulation. Lacker and Weinberg (1989) identify a class of one-period settings in which no manipulation is optimal and linear contracts obtain. Goldman and Slezak (2006) model the trade-off between effort inducement (which increases the optimal equity stake) and manipulation deterrence (which reduces it). More generally, the theory is related to dynamic models of the principal-agent problem, such as DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), He (2008a), Sannikov (2008) and Garrett and Pavan (2009), and the macroeconomic literature seeking to understand dynamic optimal incentives, such as Atkeson and Lucas (1992), Golosov, Kocherlakota and Tsyvinski (2003), Shimer and Werning (2008), Phelan and Skrzypacz (2008) and Farhi and Werning (2009). Our modeling setup builds on the multi-period framework of Edmans and Gabaix (2009) ("EG"), which allows us to derive contracts that are both attainable in closed form and "detail-neutral" – the functional form is independent of the noise distribution and agent's utility function. However, EG do not consider manipulation and restrict the CEO to consuming in the final period only. He (2008b) considers a dynamic setting in which the agent can privately save and also engage in a myopic action (similar to manipulation in this paper). He shows that the optimal wage pattern is non-decreasing over time, that sufficiently good past performance leads to permanent pay raises, and that severance pay is efficient. Our model has quite different specifications (multiplicative utility, continuous action choice and a cost function that is not restricted to being linear) which leads to a closed-form, scale-independent contract. Our analysis focuses on the state-dependent rebalancing and time-dependent vesting of the optimal contract, and its implementation via the dynamic incentive account.

In addition to its results, our paper contributes a number of methodological innovations. To our knowledge, it is the first to derive conditions on the model primitives which guarantee the validity of the first-order approach to solve a dynamic agency problem with private savings. An agency problem is a maximization problem subject to the agent's incentive constraints. The first-order approach replaces the incentive constraints against complex multi-period deviations with weaker local constraints (i.e. first-order conditions), with the hope that the solution to the relaxed problem satisfies all incentive constraints.[1] This method is often valid without private savings (hence the one-shot deviation principle), but it has proved problematic when the agent can save. The difficulties arise since the agent can engage in joint deviations to save and reduce effort, because savings provide insurance against future shocks to income and thus reduce the agent's incentives to exert effort in the future. Our method of guaranteeing the validity of the first-order approach centers around viewing the agent's total lifetime income as a function of his total disutility of effort. If this function is concave, the first-order approach is valid, since the agent's utility is concave in income. Our method of guaranteeing the validity of the first-order approach involves reparameterizing the agent's utility from being a function of consumption and effort to one of consumption and leisure. The new variable, leisure, is defined to ensure that the utility function is jointly concave in both arguments. We then linearize the agent's utility function and show that the linear utility function is jointly concave in leisure and manipulation (it is automatic that

[1] There are methods to verify the validity of the first-order approach which find the solution of the relaxed problem and verify global incentive compatibility of each individual solution numerically rather than finding conditions on primitives to find validity. For example, see Werning (2001) and Dittmann, Maug and Spalt (2008). Also, Williams (2008) derives conditions on primitives to guarantee the validity of the first-order approach, which apply to a range of dynamic contracting problems that do not involve private saving.

there is no incentive to save under linear utility). Since the actual utility function is concave, linearized utility provides an upper bound for the agent's actual utility. Thus, since there is no profitable deviation under a linear utility function, there is no profitable deviation under the actual utility function either.

The second methodological innovation allows us to solve for the optimal level of effort, rather than only the (exactly) optimal contract that implements a given effort level (the firs stage of Grossman and Hart (1983)). Following the argument of Fong and Sannikov (2009) that predictions of optimal contracting theories are important only insofar as they have sizable, rather than negligible, impact on profitability, we aim to derive a simple contract that is close to the optimal contract in terms of efficiency, rather than *the* complicated optimal contract. Our contract is *approximately* optimal under the assumption that firm value is significantly larger than the CEO's wage, which is indeed true in the vast majority of practical applications. Under this assumption, the difference in profitability between our contract and the optimal contract converges to 0 as firm's earnings become larger. The methodological innovation here is in proving that the contract is approximately optimal without deriving the optimal contract. To do so, we construct an upper bound on profit that any contract can attain, justify it using martingale methods, and show that our simple contract comes close to the upper bound. (See also He (2008b) who uses a related technique in a different setting.)

This paper is organized as follows. Section 2 presents the model setup, and Section 3 derives the optimal contract when the CEO has logarithmic utility, as this version of the model is most tractable. Section 4 shows that the key results continue to hold under general CRRA utility functions and autocorrelated noise. This section also provides a full justification of the DIA: it derives sufficient conditions that ensure that the agent will not undertake global deviations, and shows that the principal cannot improve upon implementing maximum effort. Section 5 concludes, Appendix A contains proofs, and Appendices B and C show that the model is robust to a variable marginal cost of effort and continuous time, respectively.

2 The Core Model

2.1 Assumptions

We consider a multiperiod model featuring a firm (also referred to as the principal) which employs a CEO (also referred to as the agent). The firm pays only one cash flow, a terminal dividend D_τ (also referred to as earnings) in the final period τ. In the core model, the terminal dividend is given by

$$D_\tau = X \exp\left(\sum_{s=1}^{\tau} (a_s + \eta_s)\right), \tag{1}$$

where X represents the baseline size of the firm and $a_s \in [0, \bar{a}]$ is the agent's action (also referred to as "effort"). The action a_s is broadly defined to encompass any decision that improves firm value but is personally costly to the manager. The main interpretation is effort, but it can also refer to rent extraction, in which case a low a_s reflects cash flow diversion or private benefit consumption. η_s is noise, which is independent across periods and has a log-concave density[2] with support $[\underline{\eta}, \bar{\eta}]$, where the bounds need not be finite. (Section 4.1 allows for autocorrelated noises). As in Edmans and Gabaix (2009), we assume that, in each period t, the agent privately observes η_s before choosing his action a_s. They show that this assumption leads to tractable contracts in discrete time, as well as consistent results with the continuous time case, where noise and actions are simultaneous. This timing is also featured in cash flow diversion models where the CEO sees total output before deciding how much to divert (e.g. DeMarzo and Fishman (2007)), as well as models in which the CEO observes the "state of nature" before choosing his effort level (e.g. Harris and Raviv (1979), Sappington (1983), Laffont and Tirole (1986) and Baker (1992)).

After the action is taken at time t, the principal observes a public signal of firm value,

[2] 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 (see, e.g., Caplin and Nalebuff (1991)). On the other hand, most fat-tailed distributions are not log-concave, such as the Pareto distribution.

given by:

$$S_t = X \exp\left(\sum_{s=1}^{t} (a_s + \eta_s)\right).$$

The incremental news contained in S_t, over and above the information known in period $t-1$ (and thus contained in S_{t-1}) can be summarized by $r_t = \ln S_t - \ln S_{t-1}$, i.e.

$$r_t = a_t + \eta_t. \tag{2}$$

r_t captures the additional "news" about D_τ that arises in period t. With a slight abuse of terminology, we call r_t the firm's "return" for the remainder of the paper.[3] By observing S_t, the principal learns r_t, but not its constituent components a_t and η_t. The agent's strategy is a function $a_t(r_1, \ldots r_{t-1}, \eta_t)$ that specifies how his action depends on the current level of noise for each history of returns before time t.

After S_t (and thus r_t) is publicly observed, the principal pays the agent an amount y_t according to the contract. We allow for a fully history-dependent contract in which the agent's compensation $y_t(r_1, \ldots r_t)$ in period t depends on the entire history of past returns. [4]

Having received income y_t, the agent consumes c_t and saves $(y_t - c_t)$ at the risk-free rate R. We allow $(y_t - c_t)$ to be negative, i.e. the agent may borrow as well as save. Such borrowing and saving are unobserved by the principal. Following a standard argument, we can restrict attention to contracts in which the agent chooses not to save or borrow, and

[3] r_t is the actual increase in the expected dividend as a result of the action and noise at time t. Given rational expectations, the stock return is the *unexpected* increase in firm value. In turn, firm value is the discounted expected dividend. We later show that the optimal contract implements the maximal effort $\bar{a}$ in every period. Therefore, firm value is given by

$$P_t = X \exp\left(\sum_{s=1}^{t} (a_s + \eta_s) + (\tau - t)\left(\bar{a} - R + \ln E\left[e^{\eta_t}\right]\right)\right),$$

where R is the risk-free rate. Therefore, the firm's log return is $\ln P_t - \ln P_{t-1} = R_t - \bar{a} + R$.

[4] A fully general contract can also involve the agent sending a message regarding η_t and the income y_t depending on such messages. However, such messages are redundant: the agent's announcement of η_t would be uniquely determined by r_t: since he will make the announcement that maximizes his expected utility. Therefore, the principal can automatically back out the message after seeing r_t, and so such messages would convey no additional information on top of what is already known from the history of returns. See also Edmans and Gabaix (2009).

instead consumes his entire income in each period (i.e. $c_t = y_t$). Our analysis develops a number of methodological contributions, since the analytical challenges of dynamic agency problems with private savings are well recognized.[5]

The agent's utility over consumption $c_t \in [0, \infty)$ and effort a_t in each period is given by

$$u(c_t h(a_t)), \tag{3}$$

where u is a CRRA utility function with relative risk aversion coefficient $\gamma > 0$, i.e.

$$u(x) = \begin{cases} \frac{x^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \ln x & \text{if } \gamma = 1, \end{cases}$$

and h is a decreasing, concave function.

The agent lives in periods 1 through $T \leq \tau$ and retires after period $L \leq T$. After retirement, the firm replaces him with a new CEO and continues to contract optimally. The agent discounts future utility at rate ρ, so that his total discounted utility is given by:

$$U = \sum_{t=1}^{T} \rho^t u(c_t h(a_t)). \tag{4}$$

As in Edmans, Gabaix and Landier (2009), we model effort as having a multiplicative effect on both CEO utility (equation (3)) and firm earnings (equation (1)). Multiplicative preferences consider private benefits as a normal good (i.e. the utility they provide is increasing in consumption), consistent with the treatment of most goods and services in consumer theory; they are also common in macroeconomic models. With a multiplicative production function, effort has a percentage effect on firm earnings and so the dollar benefits of working are higher for larger firms. This assumption is plausible for the majority of CEO actions, since they can be "rolled out" across the entire firm and thus have a greater effect in a larger company. Edmans et al. show that multiplicative specifications are necessary to generate empirically consistent predictions for the scaling of various measures of CEO incentives with firm size.[6]

The principal is risk-neutral and uses discount rate R, continuously compounded. There-

[5] See Werning (2002), Kocherlakota (2004) and Williams (2006).

[6] They also allow to analyze incentives in a market equilibrium model à la Gabaix and Landier (2008).

fore, her objective function is given by:

$$\max_{\{a_t,t=1,...L\},\{y_t,t=1,...T\}} E\left[e^{-R\tau}D_\tau - \sum_{t=1}^{T} e^{-Rt}y_t\right]$$

i.e. the expected discounted dividend, minus the expected cost of compensation. The individual rationality constraint is that the agent achieves his reservation utility of $\underline{u}$, i.e.

$$E^a\left[\sum_{t=1}^{T}\rho^t u(c_t h(a_t))\right] = \underline{u}.$$

The incentive compatibility constraints require that any deviation (in either the action or consumption) by the agent reduces his utility, i.e.

$$E^{\hat{a}}\left[\sum_{t=1}^{T}\rho^t u(c_t h(\hat{a}_t))\right] \leq \underline{u}$$

for all alternative effort strategies $\{\hat{a}_t, t = 1, \ldots L\}$ and *feasible* consumption strategies $\{c_t, t = 1, \ldots T\}$. A consumption strategy is feasible if it satisfies the budget constraint

$$\sum_{t=1}^{T} e^{-Rt}c_t \leq \sum_{t=1}^{T} e^{-Rt}y_t.$$

We used the notation E^a and $E^{\hat{a}}$ to highlight that the agent's effort strategy affects the probability distribution over return paths.

In some versions of the model, we allow the agent to affect the firm's returns not only by exerting effort but also via manipulation. In practice, such manipulation can take many forms. In the most literal interpretation, the manager can change accounting policies to accelerate the realization of revenues or delay the impact of costs (either by concealing information, or capitalizing rather than expensing costs).[7] Alternatively, he can engage in short-termist behavior by scrapping positive-NPV investments (as modeled by Stein (1988) and Edmans (2009)) or taking on negative-NPV projects that generate an immediate return but have a downside that may not manifest for several years (such as sub-prime lending).

[7] See Goldman and Slezak (2006) and Peng and Roell (2008, 2009) for models featuring such manipulation.

In both cases, the increase in current returns are at the expense of long-run fundamental value. Note that manipulation may be downwards as well as upwards: the CEO may sacrifice current returns to boost future returns if his future performance is benchmarked against past performance, or if the contract's sensitivity to performance increases over time. This can be achieved by investing in negative-NPV projects, or "big bath" accounting (taking large write-downs in the current period).

In each period $t \leq L$, at the same time as taking his action, the agent can also engage in manipulation $m_{t,i}$, simultaneously selecting a "release lag" $i \leq M$. The release lag is the number of periods before the effects of manipulation are reversed. For example, forgoing an investment project that pays off in the very long-run will only worsen earnings far into the future, and so the release lag is high. M is the maximum release lag, where $M \leq \tau - L$, i.e. the effects of all manipulation are reversed before the terminal dividend is paid. The terminal dividend (1) now becomes

$$D_\tau = X \exp\left(\sum_{s=1}^{\tau} (\eta_s + a_s) - \sum_{s=1}^{\tau}\sum_{i=1}^{M} \lambda(m_{s,i})\right),$$

where $\lambda(m_{s,i})$ is the fundamental cost of manipulation. We have $\lambda(0) = \lambda'(0) = 0$, and $\lambda''(m_{s,i}) > 0$. *[XG It was written*$\lambda'(|m_{s,i}|) > 0, \lambda''(|m_{s,i}|) > 0$ *but is should bejust* $\lambda''(m_{s,i}) > 0$, with no $\lambda'(|m_{s,i}|) > 0$:we don't necessarily assume that λ is symmetrical around 0, and the motonicity assumption isnt' necessary*]* Manipulation reduces fundamental value, since it involves undertaking negative-NPV projects, forsaking positive-NPV projects, or using resources to change accounting policies. For conciseness, we will sometimes shorten $m_{s,i}$ to m_s where there is no ambiguity.

The advantage of manipulation to the agent is that it temporarily increases outsiders' perceptions of firm value, and thus the firm's returns and his income. To show how manipulation affects signals and returns, we illustrate for simplicity the case where the CEO engages in only one manipulation, at time t. The signal now changes from

$$S_t = X \exp\left(\sum_{s=1}^{t} (\eta_s + a_s)\right)$$

to

$$S'_{t+j} = \begin{cases} S_{t+j} e^{m_{t,i} - \lambda(m_{t,i})} & \text{for } j = 0, ..., i-1 \\ S_{t+j} e^{-\lambda(m_{t,i})} & \text{for } j \geq i. \end{cases}$$

The return now changes from $r_t = a_t + \eta_t$ to

$$\begin{aligned} r'_t &= r_t + m_{t,i} - \lambda(m_{t,i}) \\ r'_{t+i} &= r_{t+i} - m_{t,i} \\ r'_s &= r_s & \text{for } s \neq t, t+i, \end{aligned}$$

i.e. it rises in period t by $m_{t,i} - \lambda(m_{t,i})$ and falls it in period $t+i$ by $m_{t,i}$.[8]

The principal's problem is complex because contracts are history-dependent, the agent can manipulate returns and privately save, and the principal must choose the optimal level of effort and manipulation. Our strategy for solving the problem is as follows. We start with a guess that, if the firm is sufficiently large (X is sufficiently high), we can attain the optimum at least approximately by a contract that enforces maximal effort and zero manipulation in each period, and in which the local constraints bind. Following this guess we

- characterize the class of contracts that satisfy the local incentive constraints in Section 3.1. This class includes all incentive-compatible contracts, but some contracts from the class may not be fully incentive-compatible.
- construct a candidate contract based on our guess and these characterizations in Section 3.2.
- verify that the candidate contract is also fully incentive-compatible. See Theorem 3 in

[8] If the CEO engages in multiple manipulations at time t, the signal becomes:

$$S'_{t+j} = S_{t+j} \exp(m_{s,i} 1_{s+i>t} + \sum_{\substack{s \leq t \\ i \leq M}} -\lambda(m_{s,i}))$$

and the return changes to:

$$r'_t = r_t + \sum_{i=1}^{M} (m_{t,i} - \lambda(m_{t,i})) - \sum_{i=1}^{\min\{M,t-1\}} m_{t-i,i}.$$

Section 4.2.

- verify that the candidate contract is approximately optimal among all contracts that satisfy the local incentive constraints. Specifically, Theorem 4 in Section 4.3 shows that the difference in profitability between our contract and the optimal contract converges to 0 as X becomes large.

3 Log Utility

3.1 Local Constraints

In this section we characterize the class of contracts that satisfy the local incentive constraints. There are (up to) three such constraints. The incentive compatibility (IC) constraint ensures that the agent wishes to exert the maximum level of effort ($a_t = \overline{a}$). The private savings (PS) constraint ensures that the agent wishes to consume the full income provided by the contract ($c_t = y_t$). The no-manipulation (NM) constraint ensures that the agent will not engage in manipulation ($m_t = 0$). To show the effect of allowing private savings and manipulation on the contract, we will consider versions of the model in which the PS and/or NM constraints are not imposed. We use these constraints to identify a candidate contract in Section 3.2, which we later show to be approximately optimal and fully incentive-compatible.

Consider an arbitrary contract $\{y_t, t = 1, \ldots T\}$ together with a consumption strategy $\{c_t, t = 1, \ldots T\}$, an effort strategy $\{a_t, t = 1, \ldots L\}$ and a manipulation strategy $\{(m_t, i_t), t = 1, \ldots L\}$. Recall that y_t, c_t and (m_t, i_t) depend on the entire history $(r_1, \ldots r_t)$ and a_t depends on $(r_1, \ldots r_{t-1}, \eta_t)$.[9] To capture history-dependence, we denote by E_t the expectation conditional on the history $(r_1, \ldots r_t)$.

We first address the IC constraint and consider a local deviation in the action a_t after history $(r_1, \ldots r_{t-1}, \eta_t)$. The derivative of CEO utility with respect to a_t is

$$E_t \left[\frac{\partial U}{\partial r_t} \frac{\partial r_t}{\partial a_t} + \frac{\partial U}{\partial a_t} \right],$$

[9] Since the agent has observed η_t, his action choice pins down r_t and so he knows r_t when choosing his manipulation.

where $\partial r_t/\partial a_t = 1$ and $\partial U/\partial a_t = e^{-\rho t}h'(a_t)u'(c_t h(a_t))$. The IC constraint is thus:

$$IC : E_t\left[\frac{\partial U}{\partial r_t}\right]\left\{\begin{array}{c}\geq \text{ if } a_t = \bar{a} \\ = \text{ if } a_t \in (0,\bar{a}) \\ \leq \text{ if } a_t = 0\end{array}\right\}\rho^t c_t(-h'(a_t))u'(c_t h(a_t)). \tag{5}$$

We next consider the PS constraint. If the CEO saves a small amount d_t in period t and invests it until $t+1$, his utility increases to the leading order by:

$$-E_t\left[\frac{\partial U}{\partial c_t}\right]d_t + E_t\left[\frac{\partial U}{\partial c_{t+1}}\right]e^R d_t.$$

To deter private saving or borrowing, this change should be zero to the leading order, i.e.

$$EE : \rho^t h(a_t)u'(c_t h(a_t)) = E_t\left[\rho^{t+1}e^R h(a_{t+1})u'(c_{t+1}h(a_{t+1}))\right]. \tag{6}$$

This is the standard Euler equation for consumption smoothing: discounted marginal utility $e^{Rt}\rho^t h(a_t)u'(c_t h(a_t))$ is a martingale. Intuitively, if it were not a martingale, the agent would privately reallocate consumption to the time periods with higher marginal utility.

The Euler equation can be contrasted with the "Inverse Euler Equation" (IEE), which characterizes solutions to agency problems where the agent cannot privately save and so the PS constraint need not be imposed (Rogerson (1985), Golosov, Kocherlakota and Tsyvinski (2003) and Farhi and Werning (2009)), when utility is additively separable in consumption and effort. In our model, utility becomes additive if $u(x) = \ln x$, and so the IEE is:

$$IEE\text{: } e^{-Rt}\rho^{-t}c_t \text{ is a martingale.} \tag{7}$$

The IEE states that the inverse of the agent's marginal utility, which equals the marginal cost of delivering utility to the agent, is a martingale. If (7) did not hold, the principal could benefit by shifting the agent's utility to periods with a lower marginal cost of delivering utility. This argument is invalid for $\gamma \neq 1$, because the agent's marginal cost of effort depends on his consumption when utility is nonadditive.

Finally, we consider the NM constraint. If the agent engages in a small manipulation

(m_t, i_t) at time t, his utility changes to the leading order by

$$E_t \left[\frac{\partial U}{\partial r_t} \right] (m_{t,i} - \lambda (m_{t,i})) + E_t \left[\frac{\partial U}{\partial r_{t+i}} \right] (-m_{t,i}).$$

To prevent manipulation, this change must be zero. Since $\lambda(0) = \lambda'(0) = 0$, this implies

$$NM : E_t \left[\frac{\partial U}{\partial r_t} \right] = E_t \left[\frac{\partial U}{\partial r_{t+i}} \right] \text{ for } t \le L,\ 0 \le i \le M. \tag{8}$$

We will consider versions of the model in which private savings and/or manipulation are impossible (and so the PS and/or NM constraints need not be imposed), to demonstrate how the possibility of private savings and/or manipulation affect the contract.

3.2 The Contract

We now derive the cheapest contract that satisfies the local constraints and implements maximum effort. We first present the contract under log utility, as the expressions are most transparent and the key principles are the same as in the general CRRA case. Section 4 considers the general CRRA case and extends the model to autocorrelated noise. For the derivation, it is useful to introduce the increasing function:

$$g(a) = -\ln h(a), \tag{9}$$

which represents the utility cost of exerting action a.

Theorem 1 *(Log utility.) The cheapest contract that satisfies the constraints and implements maximum effort is as follows. In each period t, the CEO's incremental utility over the previous period is linear in the signal r_t, i.e.*

$$\ln c_t = \ln c_0 + \sum_{s=1}^{t} \theta_s r_s + \sum_{s=1}^{t} k_s, \tag{10}$$

where θ_s and k_s are constants. If manipulation is impossible, the slope θ_s is given by

$$\theta_s = \begin{cases} \frac{g'(\bar{a})}{1+\rho+\ldots\rho^{T-s}} \text{ for } s \le L, \\ 0 \text{ for } s > L. \end{cases} \tag{11}$$

If manipulation is possible, θ_s *is given by:*

$$\theta_s = \begin{cases} \frac{g'(\bar{a})}{1+\rho+\ldots\rho^{T-s}}\rho^{1-s} \text{ for } s \leq L+M, \\ 0 \text{ for } s > L+M \end{cases} \tag{12}$$

If private saving is impossible, the constant k_s *is given by:*

$$k_s = R + \ln \rho - \ln E[e^{\theta_s(\bar{a}+\eta)}]. \tag{13}$$

If private saving is possible, k_s *is given by:*

$$k_s = R + \ln \rho + \ln E[e^{-\theta_s(\bar{a}+\eta)}]. \tag{14}$$

The initial condition c_0 *is chosen to give the agent his reservation utility* $\underline{u}$.

Heuristic proof. The Appendix contains a full proof; here we present a heuristic proof in a simple case, that gives the key intuition behind the formal proof. We consider a two-period model with no discounting, i.e. $L = T = 2$, $\rho = 1$, $R = 0$, with the PS constraint but without the NM constraint. We wish to show that the optimal contract is written:

$$\ln c_1 = g'(\bar{a}) \frac{r_1}{2} + \kappa_1, \quad \ln c_2 = g'(\bar{a}) \left(\frac{r_1}{2} + r_2\right) + \kappa_1 + k_2 \tag{15}$$

for some constants κ_1 (the equivalent of $\ln c_0 + k_1$ in the Theorem) and k_2 that makes the IR constraint bind.

Step 1: Optimal log-linear contract

We first solve the problem in a restricted class where contracts are log-linear, i.e.:

$$\ln c_1 = \theta_1 r_1 + \kappa_1, \ \ln c_2 = \theta_{21} r_1 + \theta_2 r_2 + \kappa_1 + k_2 \tag{16}$$

for some constants θ_1, θ_{21}, θ_2, κ_1, k_2 to be determined. This first step is not necessary, but it clarifies the economics, and it is helpful in more complicated cases in guessing the form of the optimal contract.

First, intuitively, the optimal contract should entail consumption smoothing, i.e. a shock to consumption has a permanent impact. This implies $\theta_{21} = \theta_1$. To see this more formally, the Euler Equation (6) yields:

$$1 = E_1 \left[\frac{c_1}{c_2} \right] = e^{(\theta_1 - \theta_{21}) r_1} E_1 \left[e^{-\theta_2 r_2 - k_2} \right]. \tag{17}$$

This must hold for all r_1. Therefore, $\theta_{21} = \theta_2$.

Next, consider total utility U:

$$\begin{aligned} U &= \ln c_1 + \ln c_2 - g(a_1) - g(a_2) \\ &= 2\theta_1 r_1 + \theta_2 r_2 - g(a_1) - g(a_2) \end{aligned}$$

From (5), the two IC conditions are $E_2 \left[\frac{\partial U}{\partial r_1} \right] \geq g'(\overline{a})$ and $E_2 \left[\frac{\partial U}{\partial r_2} \right] \geq g'(\overline{a})$. They are equivalent to:

$$2\theta_1 \geq g'(\overline{a}), \quad \theta_2 \geq g'(\overline{a}).$$

It is intuitive that the IC constraints should bind, otherwise the CEO is exposed to unnecessary risk. We therefore have:

$$2\theta_1 = g'(\overline{a}), \quad \theta_2 = g'(\overline{a}).$$

Combining this equation with (16), we see that the optimal contract is given by (15).

Finally, we revisit equation (17), which gives $k_2 = \ln E_1 \left[e^{-\theta_2 r_2} \right]$, as in (13).

Step 2: Optimality of log-linear contracts

We next verify that optimal contracts should be log-linear. To do so, we use the following reasoning from EG. (5) yields: $d(\ln c_2)/dr_2 \geq g'(\overline{a})$. In the Appendix we show that the cheapest contract involves this local IC condition binding, i.e.

$$d(\ln c_2)/dr_2 = g'(\overline{a}) \equiv \theta_2. \tag{18}$$

Integrating yields the contract:

$$\ln c_2 = \theta_2 r_2 + B(r_1), \tag{19}$$

where $B(r_1)$ is a function of r_1 which we will determine shortly. It is the integration "constant" of equation (18) viewed from time 2.

We next apply the Euler Equation (6) for $t = 1$:

$$1 = E_1\left[\frac{c_1}{c_2}\right] = E_1\left[\frac{c_1}{e^{\theta_2 r_2 + B(r_1)}}\right] = E_1\left[e^{-\theta_2 r_2}\right] c_1 e^{-B(r_1)}. \tag{20}$$

Hence, we obtain

$$\ln c_1 = B(r_1) + K, \tag{21}$$

where the constant K is independent of r_1. (In this proof, expressions such as K and K' are constants independent of r_1 and r_2.) Total utility is:

$$U = \ln c_1 + \ln c_2 + K' = 2B(r_1) + K'. \tag{22}$$

We next apply (5) to (22) to yield: $2B'(r_1) \geq g'(\overline{a})$. Again, the cheapest contract involves this condition binding, i.e. $2B'(r_1) = g'(\overline{a})$. Integrating yields:

$$B(r_1) = g'(\overline{a})\frac{r_1}{2} + K'', \tag{23}$$

Combining (23) with (21) yields:

$$\ln c_1 = g'(\overline{a})\, r_1 + \kappa_1,$$

for another constant κ_1. Combining (23) with (19) yields:

$$\ln c_2 = g'(\overline{a})\left(\frac{r_1}{2} + r_2\right) + \kappa_1 + k_2,$$

for some constant k_2. ■

We now discuss the economics behind the contract. (10) shows that time-t income should be linked to the return not only in period t, but also in all previous periods. Therefore, exerting effort in a particular period boosts income in both the current and all future periods. We call this the *"deferred reward principle"*: since the CEO is risk-averse, it is optimal to spread the reward for effort across all future periods rather than concentrate it in the period in which effort is exerted. This prediction is consistent with Boschen and Smith (1995), who find that changes in firm value have a much greater effect on future rather than contemporaneous pay.

We now consider how the contract sensitivity changes over time. We first consider the case where manipulation is impossible and so the NM constraint is not imposed. (11) shows that, in an infinite horizon model ($T = \tau = \infty$), the sensitivity is constant and given by:

$$\theta_t = \theta = (1 - \rho)\, g'(\overline{a}). \tag{24}$$

This time-independent sensitivity is intuitive: the contract must be sufficiently sharp to compensate for the disutility of effort, which is constant. However, in a finite model, (11) shows that θ_t is increasing over time. The intuition for this *"increasing incentives principle"* is that there are fewer remaining periods over which to smooth out the reward for effort, and so the CEO must earn a greater reward in each period. As in Gibbons and Murphy (1992), our model predicts that CEOs closer to retirement must have sharper contracts. While Gibbons and Murphy obtain this result by invoking career concerns, we derive this result in the absence of career concerns: instead it arises because consumption smoothing possibilities decline towards retirement.

Next, we study the impact of manipulation on the contract. From (12), the possibility of manipulation has three main effects. First, it requires that the CEO's income remains sensitive to firm returns after his retirement in period L: it remains sensitive until period $L+M$, by which time all manipulation has been reversed. This is to deter him from inflating returns just before retirement. Second, it causes the contract sensitivity to be higher in each period, because the contract must now satisfy the NM constraint as well as the IC constraint. Third, it affects how the contract sensitivity trends over time. If this sensitivity were constant, the CEO would have an incentive to inflate the time-t return, thus increasing his time-t consumption. Even though the return at time $t + i_t$ will be lower, the effect on the CEO's utility is smaller owing to discounting. Therefore, an increasing sensitivity is necessary to deter manipulation. For example, in an infinite horizon model ($T = \infty$), the possibility of manipulation changes the slope from the constant (24) to

$$\theta_t = (1-\rho)\,\rho^{1-t} g'(\overline{a})$$

The ρ^{1-t} term demonstrates the increasing slope. The more impatient the CEO, the greater the incentives to manipulate, and so the greater the required increase in sensitivity over time to deter manipulation. In a finite horizon model, the slope is already increasing if manipulation is impossible; the feasibility of manipulation causes it to rise even faster.

Finally, the possibility of private savings affects the constant k_t but not the sensitivity θ_t. Since private saving does not affect the agent's action and thus firm returns, the sensitivity of CEO pay to returns is unchanged. Instead, it alters the time trend in the level of pay. The constant k_t in (14), where private savings is possible, declines more slowly (or increases more rapidly) over time than in (13) where private savings are infeasible. The faster upward trend means that the contract effectively saves for the agent, removing the need for him to do so himself. This result is consistent with He (2008b), who finds that the optimal contract under private savings involves a wage pattern that is non-decreasing over time.

The contract in Theorem 1 involves binding local constraints and implements maximum effort and zero manipulation in each period. The remaining steps are to show that the agent will not wish to undertake global deviations (e.g. make large changes, or simultaneously reduce effort, save and/or manipulate) and that the principal cannot improve significantly by implementing a different effort or manipulation level, or allowing slack constraints. Since these proofs are equally clear for general γ as for log utility, we delay them until Section 4 where we extend the model to general CRRA utility. For now, we present the intuition for the second result, that the principal cannot improve on the contract. The local constraints bind because slack constraints would increase the sensitivity of the contract, subjecting the CEO to unnecessary risk. The optimal effort level is the result of a trade-off between the benefits of effort (which are multiplicative in firm size) and the costs of effort. The latter are the direct disutility suffered by the CEO from working, plus the inefficient risk-sharing caused by giving the CEO variable consumption, and are a function of the CEO's salary. If firm size is sufficiently large compared to the CEO's salary, the benefits of effort swamp the costs and maximum effort is optimal. A similar argument applies to manipulation.

Appendix B extends the analysis to a variable marginal cost of effort; the key results are unchanged.

3.2.1 A Numerical Example

This optional section uses a simple numerical example to show most clearly the deferred reward and increasing incentives principles, as well as the effect of manipulation on the contract. We first set $T = 3$, $L = 3$, $\rho = 0$ and $g'(\overline{a}) = 1$, and assume that manipulation is impossible. From (11), the contract is given by:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{3} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{3} + \frac{r_2}{2} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{3} + \frac{r_2}{2} + \frac{r_3}{1} + \kappa_3
\end{aligned}$$

where the $\kappa_t = \sum_{s=1}^{t} k_s$ are constants. This example shows both principles at work. First, there is consumption smoothing: an increase in r_1 leads to a permanent increase in log consumption (and thus utility) – it rises by $\frac{r_1}{3}$ in all future periods. Second, the sensitivity increases over time, from 1/3 to 1/2 to 1/1.

We now allow the CEO to continue to live after he retires, by now considering $T = 5$ but retaining all of the previous parameters. The optimal contract is now:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3 \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_4 \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_5.
\end{aligned}$$

Since the CEO takes no action from $t = 4$ onwards, his pay does not depend on r_4 or r_5. However, it continues to depend on r_1, r_2 and r_3 as his earlier efforts affect his wealth, from which he consumes until death.

If the CEO can manipulate returns with $M = 1$, the contract changes to:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3 \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{r_4}{2} + \kappa_4 \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{r_4}{2} + \kappa_5.
\end{aligned}$$

The possibility of manipulation means that r_4 now affects the CEO's income, otherwise he would have an incentive to boost r_3 at the expense of r_4. However, the contract is unchanged for $t \leq 3$, i.e. for the periods in which the CEO works. Even under the original contract, there is no incentive to manipulate at $t = 1$ or $t = 2$ because two conditions are satisfied. First, there is no discounting, and so the negative effect of manipulation on future returns reduces the CEO's lifetime utility by as much as the positive effect on current salary increases it. Comparing (11) and (12) shows that, if $\rho < 1$ (i.e. there is discounting), the possibility of manipulation causes the contract slope to rise at all t. Second, because the marginal cost of effort is constant across periods, the lifetime effect of increasing returns is the same regardless of the period in which the higher returns arise. For example, increasing r_1 by one unit raises consumption in each period by 1/5 units, and so 1 unit (undiscounted) in total. Decreasing r_2 by one unit reduces consumption in each period by 1/4 units, and so 1 unit in total. Again, the costs and benefits of manipulation are the same, so there is no incentive to manipulate even under the original contract.

3.3 Implementation of the Contract: the DIA

From Theorem 1, we have

$$\ln c_t - \ln c_{t-1} = \theta_t r_t + k_t. \tag{25}$$

The contract thus prescribes the percentage change in CEO pay as a function of the firm's return r_t, i.e. the percentage change in firm value. The relevant measure of incentives is therefore the elasticity of CEO pay to firm value; this elasticity must be at least θ_t to ensure incentive compatibility. Empiricists have used a number of statistics to measure incentives – for example, Jensen and Murphy (1990) calculate "dollar-dollar" incentives (the dollar change in CEO pay for a dollar change in firm value) and Hall and Liebman (1998) measure "dollar-percent" incentives (the dollar change in CEO pay for a percentage firm return.) By contrast, Murphy (1999) advocates the elasticity measure ("percent-percent" incentives) on 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 provides a theoretical justification for using elasticities to measure incentives. Edmans, Gabaix and Landier (2009) showed that percent-percent incentives are the optimal measure if effort has a multiplicative effect on both CEO utility and firm value, as in this paper (equations (1) and (3)).[10] Their result was derived in a one-period model with a risk-neutral CEO; we extend it to a dynamic model with a risk-averse CEO who can manipulate returns and privately save. In terms of real variables, percent-percent incentives equal the fraction of total pay that is comprised of stock. The required fraction (θ_t) is independent of total pay (i.e. scale-independent): if the CEO's outside option doubles, total pay doubles. Therefore, the value of equity must double to ensure that the fraction of total pay invested in equity remains the same.

To ensure that percent-percent incentives equal θ_t in each period t, the contract can be implemented in the following simple manner. The present value of the CEO's expected pay is escrowed into a "Dynamic Incentive Account" ("DIA") at the start of period $t = 1$.[11] A proportion θ_1 of the Incentive Account is invested in the firm's stock and the remainder in cash.[12] At the start of each subsequent period t, this portfolio is rebalanced so that the proportion invested in the firm's stock is θ_t. This dynamic rebalancing addresses a common problem of option compensation: if firm value declines, the option's delta falls and

[10] "Percent-percent" incentives are also the optimal measure in Peng and Roell (2008).

[11] We present one possible implementation of the optimal contract; other implementations are possible. For example, rather than placing the entire present value of the CEO's future pay in the account at the start, only his $t = 1$ reservation wage could be invested initially. In each subsequent period, the reservation wage of that period is added to the account.

[12] Note that the stock pays the firm's actual return. As noted in footnote 3, r_t is not the firm's actual return, but the actual return plus $\overline{a} - R$. This does not affect the implementabiity with stock because it only changes the constant k_t, which rises by $\theta_t(\overline{a} - R)$.

so its incentive effect is reduced. Unrebalanced stock compensation suffers from a similar problem, even though the delta of a share is constant at 1 regardless of firm value. The relevant measure of incentives is not the delta of the CEO's portfolio (which represents dollar-dollar incentives) but the proportion of CEO pay which is in equity (percent-percent incentives). When the stock price falls, the value of the CEO's shares declines but his cash is unaffected. Therefore, stock constitutes a smaller proportion of the CEO's pay, which reduces his incentives. The DIA addresses this problem by exchanging cash for stock, to maintain the fraction of stock in the account at θ_t. Importantly, the additional stock is accompanied by a reduction in cash – it is not given for free. This addresses a major concern with repricing options after negative returns to restore incentives – the CEO is rewarded for failure. By contrast, if the stock price rises, stock becomes a higher fraction of the account. Therefore, some shares can be sold for cash, thus reducing the CEO's risk, without incentives falling below θ_t. Indeed, Fahlenbrach and Stulz (2008) find that decreases in CEO ownership typically occur after good performance.

The DIA thus features dynamic rebalancing to ensure that the IC constraint is satisfied in each period. This rebalancing is state-dependent: if the stock price rises (falls), stock is sold (bought) for cash. The second key feature of the DIA is time-dependent vesting: the CEO can only withdraw a fraction α_t of the account in each period for consumption (we will later derive α_t in a specific case). This gradual vesting ensures that the NM constraint is satisfied in each period: it prevents the CEO from manipulating returns and then cashing out his equity before the manipulation is revealed. Moreover, vesting is gradual not only during the CEO's tenure but also after retirement. The CEO is not paid the entire DIA in period L. Instead, the account only fully vests in period $L + M$, to deter the CEO from inflating returns just before his departure. Commentators have argued that the latter problem was particularly important in the recent financial crisis. For example, Angelo Mozilo, the former CEO of Countrywide Financial, made over $100m from stock sales prior to his firm's collapse; a November 20, 2008 *Wall Street Journal* article entitled "Before the Bust, These CEOs Took Money Off the Table" provides further examples. More broadly, Johnson, Ryan and Tian (2009) find a positive correlation between corporate fraud and unrestricted (i.e. immediately vesting) stock compensation.

In sum, the DIA has two key features. Time-dependent vesting ensures that the CEO does not manipulate returns, while smoothing his consumption so that he has no incentive to privately save. State-dependent rebalancing guarantees that the CEO has sufficient incentives to exert effort, while minimizing the risk that he bears. Some existing compensation schemes satisfy the first feature, but not the second. For example, restricted stock and options vest along a given time schedule, irrespective of firm performance (see, e.g., Kole (1997)). Long-vesting securities are effective in satisfying the NM constraint but not the IC constraint when firm value changes over time. Hence, the DIA is critically different from the restricted securities observed empirically.

Time-dependent vesting is not the only schedule seen in practice. Bettis, Bizjak, Coles and Kalpathy (2008) show that performance-based (i.e. state-dependent) vesting is becoming increasingly common. State-dependent vesting is also featured in the "Bonus Bank" advocated by Stern Stewart, where the amount of the bonus that the executive can withdraw depends on the total bonuses accumulated in the bank. Under performance vesting, the vesting schedule is accelerated if the firm performs strongly. This may induce the CEO to inflate returns to accelerate vesting, and sell his equity before the manipulation is reversed. In the DIA, strong performance allows the CEO to sell his shares for cash, but critically the cash is maintained within the DIA to allow for future stock repurchases if the stock price later falls. The combination of time-dependent vesting and state-dependent rebalancing thus achieves a different result from state-dependent vesting – the two separate features achieve the two goals of deterring manipulation and maintaining effort incentives.

We demonstrate the workings of the DIA in an infinite horizon model ($T = \infty$) were manipulation is impossible. The contract sensitivity is constant and given by (24). The CEO's consumption is:

$$c_t = c_0 e^{\theta R_t + nt}, \text{ where } n \equiv R + \ln \rho + \ln E\left[e^{-\theta(\overline{a}+\eta)}\right]. \tag{26}$$

To obtain easy-to-interpret closed forms, we take the continuous time limit of the problem. Let $A_0 = E_0\left[\int_0^T e^{-Rt} c_t dt\right]$ be the initial value of the DIA, i.e. the present value of future consumption under maximum effort and no manipulation. A fraction θ is invested in the firm's stock and the remainder in cash. This fraction is continuously rebalanced so that the account evolves according to: $dA_t/A_t = (R - \alpha)\, dt + \theta\sigma dZ_t$. The CEO withdraws a fraction α of the account in each period, so that his consumption is $c_t = \alpha A_t$. This is intuitive, since an agent with log utility wishes to consume a constant fraction of his wealth in each period, and this fraction is independent of the return on his wealth. When the agent retires, he receives the entire remaining value of the DIA.

If the PS constraint is not imposed, from (26) we have $\alpha = -\ln \rho$ and the inverse marginal utility, c_t, is a martingale. If the PS constraint is imposed, $\alpha = -\ln \rho - \sigma^2\theta^2 < -\ln \rho$. The intuition is as follows. The agent would like to invest zero wealth in the stock as it carries a zero risk premium, but he is forced to invest θ and bear unrewarded risk. Therefore, the agent will wish to save to insure himself against this risk. To remove these incentives, we must have $\alpha < -\ln \rho$ so that the account grows faster than it vests, thus providing automatic saving for the agent.

4 Generalization and Justification

Section 4.1 generalizes our contract to all CRRA utility functions and autocorrelated noise. Section 4.2 verifies that the candidate contract is fully incentive compatible (i.e. the agent does not wish to undertake global deviations) and Section 4.3 proves that the candidate contract is approximately optimal. Specifically, we show that as firm size increases, the difference in profit between our contract (involving maximum effort, zero manipulation and binding constraints) and the possibly much more complicated optimal contract goes to 0.

4.1 General CRRA Utility and Autocorrelated Signals

The core model assumes that the signal r_t was the firm's stock return. This is an attractive interpretation for a number of reasons: it allows the optimal contract to be implemented using the firm's securities, and it allows us to assume that the noises η_t are uncorrelated. However, in private firms, there is no stock return, and so alternative signals of effort must be used such as profits. Unlike stock returns, shocks to profits may be serially correlated. This subsection extends the model to such a case. We now assume that the noises $\eta_1, ..., \eta_T$ follow an $AR(1)$ process with autoregressive parameter ϕ, i.e. $\eta_t = \phi\eta_{t-1} + \varepsilon_t$, $\phi \in [0, 1]$, where ε_t are independent with support $(\underline{\varepsilon_t}, \overline{\varepsilon_t})$; the bounds need not be finite.

We also now allow for a general CRRA utility function. Note that for $\gamma \neq 1$, the IEE is not valid for the case where private savings are impossible, so we only consider the case where the PS constraint is imposed. Define $B_t = \rho^t e^{-(1-\gamma)g(\overline{a})}$.

Theorem 2 *(General CRRA utility, autocorrelated noise, with Private Savings constraint) The cheapest contract that satisfies the constraints and implements maximum effort is as follows. In each period t, the CEO is paid c_t which satisfies:*

$$\ln c_t = \ln c_0 + \sum_{s=1}^{t} \theta_s \left(r_s - \phi r_{s-1}\right) + \sum_{s=1}^{t} k_s, \tag{27}$$

where θ_s and k_s are constants, and $r_0 = 0$. If manipulation is impossible, the slope θ_s is given by:

$$\theta_s = \begin{cases} \frac{B_t(g'(\overline{a}) - \phi\theta_{t+1})}{\sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + \overline{a}(1-\phi)) + k_n]}\right]} + \phi\theta_{t+1} \text{ for } t \leq L, \\ 0 \text{ for } t > L. \end{cases} \tag{28}$$

If manipulation is possible, θ_s is given by:

$$\theta_t = 0 \text{ for } t > L + M,$$
$$\theta_t = \frac{D \prod_{n=t+1}^{L+M} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + \overline{a}(1-\phi)) + k_n]}\right] - B_t\phi\theta_{t+1}}{\sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + \overline{a}(1-\phi)) + k_n]}\right]} + \phi\theta_{t+1} \text{ for } t \leq L + M.$$

The constant k_t is given by

$$\gamma k_t = r + \ln\rho + \ln E\left[e^{-\gamma\theta_t(\varepsilon_t + \overline{a}(1-\phi))}\right] \text{ for } t \leq T. \tag{29}$$

The initial condition c_0 is chosen to give the agent his reservation utility $\underline{u}$, and D is the lowest constant such that:

$$D \prod_{n=t+1}^{L+M} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + \overline{a}(1-\phi)) + k_n]}\right] \geq B_t g'(\overline{a}), \text{ for all } t \leq L.$$

Proof See Appendix. ■

From (27) we can see the effect of allowing for general CRRA utility functions and autocorrelated noise. With independent noise $\phi = 0$ and so contracts (27) reduces to (10) and (25). Therefore, moving from log to CRRA utility but retaining independent noise has little effect on the functional form of the optimal contract. The deferred reward and increasing incentive principles, the effect of the NM constraint, and the implementation via the DIA remain the same. The difference is that the parameters θ and k are somewhat more complex.

Equation (27) shows that, in the presence of autocorrelated signals, the optimal contract now links the percentage change in CEO pay in period t to innovations in the signal $(r_t - \phi r_{t-1})$ between t and $t-1$, rather than the absolute signal in period t. This is intuitive: since good luck (i.e. a positive shock) in the last period carries over to the current period, the contract should control for the last period's signal to avoid paying the CEO for luck.

Appendix C analyzes a further extension, to continuous time. The contract is consistent with the discrete time case.

4.2 Global Constraints

We have thus far analyzed the first stage of the derivation of the optimal contract, which is to find the best contract that satisfies the local constraints. The second stage is to verify that this contract also satisfies the global constraints, i.e. the agent does not wish to undertake global deviations (large changes, or jointly shirking, saving and/or manipulating). At present, the analysis assumes either $\gamma = 1$ or $\phi = 0$. It will be generalized in a later draft.

The contract in Theorem 2 pays the agent an income y_t, given by:

$$\ln y_t = \sum_{s=1}^{t} \theta_s (r_s + \overline{m}_s - \phi(r_{s-1} + \overline{m}_{s-1})) + k_t, \tag{30}$$

where

$$\overline{m}_s = \sum_{i=1}^{M} (m_{s,i} - \lambda(m_{s,i})) - \sum_{i=1}^{\min\{M,s-1\}} m_{s-i,i} \tag{31}$$

is the overall effect of manipulations on the return in period s.

The following Theorem states that if the cost functions g and λ are sufficiently convex, the CEO has no profitable global deviation.

Theorem 3 *(No global deviations are profitable.) Consider the maximization problem:*

$$\max_{a_t, c_t, m_t \text{ adapted}} E\left[\sum_{t=1}^{T} \rho^t \frac{\left(c_t e^{-g(a_t)}\right)^{1-\gamma}}{1-\gamma}\right], \text{ for } \gamma \neq 1 \tag{32a}$$

$$\max_{a_t, c_t, m_t \text{ adapted}} E\left[\sum_{t=1}^{T} \rho^t (\ln c_t - g(a_t))\right], \text{ for } \gamma = 1, \tag{32b}$$

with $\sum_{t=1}^{T} e^{-rt}(y_t - c_t) \geq 0$ *and* y_t *satisfying (30). Assume* $\gamma = 1$ *or* $\phi = 0$*. If functions* g *and* λ *are sufficiently convex, i.e.* $\inf_m \lambda''(m)$ *and* $\inf_a g''(a)$ *are sufficiently large, the solution of this problem is* $c_t \equiv y_t$, $t \leq T$, *and* $a_t \equiv \overline{a}$, $m_t = (m_{t,1}, ..., m_{t,M}) \equiv \mathbf{0}$, $t \leq L$*. In other words, there is no global deviation from the recommended policy that makes the agent better off.*

The proof, in the Appendix, may be of general methodological interest. It involves three key steps. First, we reparameterize the agent's utility from being a function of consumption and effort to one of consumption and leisure, where the new variable, leisure, is defined to ensure that the utility function is jointly concave in both arguments. Second, we construct an "upper-linearization" function: we create a surrogate agent with a linear state-dependent utility. Since the original agent's utility function is concave, the linear utility function is always weakly higher than the utility of our original agent, and the same at the recommended policy. Third, we prove that any global deviation by the surrogate agent weakly reduces his utility below that under the recommended policy. As the surrogate agent's utility is linear, it is automatic that there is no motive to save; we then show that the present value of the agent's income is concave in the agent's two other decisions, leisure (and thus effort) and manipulation. Since consumption equals income, and utility is linear in consumption, the utility function is concave in leisure and manipulation and so there is no profitable deviation. Since our original agent's utility is the same as the surrogate agent's under the recommended policy, and weakly lower under any other policy, any deviation to another policy also reduces the original agent's utility. *[AE: we used to say "The second argument is a potentially useful Lemma that show that the present value of income is a concave function of actions under suitable reparametrization." Is this indeed the most novel part of the proof? The idea of showing that something is concave seems rather standard; to me it seems the* *reparameterization of the utility function (to concave) and then the linearization are the most novel steps that are of most methodological interest. XG Alex and I just talked about that on the phone]*

4.3 The Optimality of Maximum Effort

Edmans and Gabaix (2009) show in a discrete time, one-period setting, that if the firm is sufficiently large, it is optimal for the principal to implement the maximum effort level. This section extends this maximum effort principle to broader settings. The analysis is still in progress; this section contains the results obtained thus far. We consider a continuous-time model with a continuous dividend.

Theorem 4 *(Maximum effort is approximately optimal.) Fix* $\underline{u}$*. For any* $\varepsilon > 0$ *there exists* X_* *large enough such, if* $X > X_*$*, the principal's profit from the contract in Theorem 1 differs from his profit from the optimal contract by at most* ε*.*

Hence, we show that contract requiring maximum effort is optimal, within an ε. The proof is will be made available soon. We suspect that an analogous, and stronger result, might be available in discrete time. In addition, we suspect that an analogous result is available to show that zero manipulation is optimal. We are currently researching these issues.

5 Conclusion

This paper studies optimal CEO compensation in a dynamic setting in which the CEO consumes in each period, can privately save, and may temporarily manipulate returns. The optimal contract involves consumption smoothing, where current effort is rewarded in all future periods, and the relevant measure of incentives is the percentage change in pay for a percentage change in firm value. This required elasticity is constant over time in an infinite horizon model where manipulation is impossible. If the horizon is finite, the contract's slope rises over time since, as the CEO approaches retirement, he has fewer periods over which to be rewarded for effort. A rising slope also arises if the contract needs to prevent manipulation. This is to offset the fact that the cost of manipulation is suffered only in the future and thus has a discounted effect on the CEO's utility. Deterring manipulation also requires the CEO to remain sensitive to the firm's stock price after retirement. While the possibility of manipulation affects the elasticity of pay to firm value, the option to privately save impacts the time trend in total pay. It augments the rise in compensation over time, removing the need for the CEO to save himself.

The optimal contract can be implemented using a Dynamic Incentive Account. The CEO's expected pay is placed into an account, and a certain proportion is invested in the firm's stock, with the remainder in cash. The account features both state-dependent rebalancing and time-dependent vesting. As firm value changes, the account is constantly rebalanced so that the proportion invested in the stock remains at the required threshold. This ensures that the CEO has adequate incentives even if the stock price falls. The gradual vesting of the account, even after retirement, allows the CEO to consume while simultaneously deterring myopic actions.

Our key results are robust to a broad range of settings: general CRRA utility functions, all noise distributions with interval support, autocorrelated noise, and continuous time. However, our setup imposes some limitations, in particular that the CEO remains with the firm for a fixed period. It would be interesting to examine how the optimal contract changes if firings and voluntary departures are possible. For example, if the CEO's outside option is stochastic, he may leave mid-way through the contract. Conversely, if the CEO becomes wealthy, his utility from shirking rises, given multiplicative preferences. This increases the cost of providing incentives and may induce the principal to replace the CEO. We leave such extensions to future research.

A Proofs

A.1 Proof of Theorem 1

The heuristic proof gave the "essence" of the argument. The formal proof is a direct corollary of Theorem 2.

A.2 Proof of Theorem 2

We first analyze the core model where manipulation is impossible. We consider the NM constraint at the end of the proof.

Case $t > L$. For $t > L$, r_t is independent of the CEO's actions. Since the CEO is strictly risk averse, c_t will depend only on $r_1, ..., r_L$. Therefore either the PS constraint (6) or IEE (if $\gamma = 1$) immediately give

$$\ln c_t(r_1, ..., r_t) = \ln c_L(r_1, ..., r_L) + \kappa_t^L, \tag{33}$$

for some constants κ_t^L independent of $r_1, r_2, ...$ that will be computed explicitly at the end of the proof.

Case $t = L$. The IC in period L requires that

$$0 \in \arg\max_{\varepsilon \leq 0} U(r_1, ..., r_{L-1}, \overline{a} + \eta_L + \varepsilon). \tag{34}$$

Since g is differentiable, this yields (5) (see e.g. EG, Lemma 6), i.e.

$$\frac{d}{d\varepsilon_-} \ln c_L\left(r_1, ..., \overline{a} + \eta_L + \varepsilon\right)|_{\varepsilon=0} \left[\sum_{s=L}^{T} B_t\right] \geq B_L g'(\overline{a}), \text{ for } \gamma = 1,$$

$$\frac{d}{d\varepsilon_-} \frac{c_L\left(r_1, ..., \overline{a} + \eta_L + \varepsilon\right)^{1-\gamma}}{1-\gamma}|_{\varepsilon=0} \left[\sum_{s=L}^{T} B_t e^{(1-\gamma)\kappa_t^L}\right] \geq B_L c_L\left(r_1, ..., \overline{a} + \eta_L + \varepsilon\right)^{1-\gamma} g'(\overline{a}), \text{ for } \gamma \neq 1.$$

and so

$$\frac{d}{d\varepsilon_-} \ln c_L\left(r_1, ..., \overline{a} + \eta_L + \varepsilon\right) \geq \frac{B_t g'(\overline{a})}{\sum_{t=L}^{T} B_t e^{(1-\gamma)\kappa_t^L}} := \theta_L. \tag{35}$$

We now show that (35) holds with equality. First, condition (35) implies that for any

$r' \geq r$ (see EG, Lemma 4)

$$\ln c_L\left(r_1, ...r_{L-1}, r'\right) - \ln c_L\left(r_1, ...r_{L-1}, r\right) \geq \theta_L(r' - r). \tag{36}$$

Consider now the contract $\{c_t^0\}_{t\leq T}$ that coincides with $\{c_t\}_{t\leq T}$ for $t < L$, $\ln c_t^0 = \ln c_L^0 + \kappa_t^L$ for $t > L$ and k_t^L as in (33), and such that $c_L^0(r_1, ..., r_L) = e^{B(r_1,...,r_{L-1})+\theta_L r_L}$, where $B(r_1, ..., r_{L-1})$ is chosen to satisfy

$$E_{L-1}\left[\frac{\left(c_L^0\right)^{1-\gamma}(r_1, ..., r_L)}{1-\gamma}\right] = E_{L-1}\left[\frac{\left(c_L\right)^{1-\gamma}(r_1, ..., r_L)}{1-\gamma}\right]. \tag{37}$$

Note that the condition (36) guarantees that the random variable $\ln c_L\left(r_1, ...r_{L-1}, \widetilde{r}_L\right)$ is weakly more dispersed than $\ln c_L^0\left(r_1, ...r_{L-1}, \widetilde{r}_L\right)$.[13] It also follows from the IC that both $\ln c_L\left(r_1, ...r_{L-1}, \cdot\right)$ and $\ln c_L^0\left(r_1, ...r_{L-1}, \cdot\right)$ are weakly increasing. Those facts together with (37) imply that for the convex function ψ and increasing function ξ, where $\psi^{-1}(x) = \frac{x^{1-\gamma}}{1-\gamma}$ and $\xi(x) = \frac{e^{(1-\gamma)x}}{1-\gamma}$, we have (see EG, Lemmas 1 and 2):

$$E_{L-1}[c_L^0(r_1, ..., r_L)] = E_{L-1}[\psi\circ\xi\circ\ln c_L^0(r_1, ..., r_L)] \leq E_{L-1}\left[\psi \circ \xi \circ \ln c_L(r_1, ..., r_L)\right] = E_{L-1}[c_L(r_1, ..., r_L).$$

Consequently the contract $\{c_t^0\}_{t\leq T}$ is cheaper than $\{c_t\}_{t\leq T}$.

Integrating out (35) that holds with equality, the optimal contract c is given by:

$$\ln c_L(r_1, ..., r_L) = B(r_1, ..., r_{L-1}) + \theta_L r_L + \kappa_L,$$

for some function B.

Case $t < L$. Suppose that for all t', $L \geq t' > t$, the optimal contract $c_{t'}$ is such that

$$\ln c_{t'}(r_1, ..., r_{t'}) = B(r_1, ..., r_t) + \theta_{t'} r_{t'} + \sum_{s=t+1}^{t'-1}(\theta_s - \phi\theta_{s+1})r_s + \kappa_{t'},$$

[13] Let X and Y denote two random variables with cumulative distribution functions F and G and corresponding right continuous inverses F^{-1} and G^{-1}. X is said to be less dispersed than Y if and only if $F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha)$ whenever $0 < \alpha \leq \beta < 1$.

for some function B as well as θ_s as in the Theorem. The PS constraint yields

$$c_t^{-\gamma} = e^R \frac{B_{t+1}}{B_t} E_t\left[c_{t+1}^{-\gamma}\right] = E_t\left[e^{-\gamma\theta_{t+1}r_{t+1}}\right] e^{-\gamma B(r_1,\ldots,r_t)+R-\gamma\kappa_{t+1}+\ln B_{t+1}-\ln B_t}. \tag{38}$$

We therefore have[14]

$$\ln c_t = B(r_1, ..., r_t) + \phi\theta_{t+1}r_t + \kappa_t. \tag{39}$$

As in the case $t = L$, the IC implies that:

$$B_t c_t^{1-\gamma}\phi\theta_{t+1} + \frac{d}{d\varepsilon_-}B\left(r_1, ...r_{t-1}, \overline{a} + \eta_t + \varepsilon\right)\sum_{s=t}^{T} B_s E_t\left(c_s^{1-\gamma}\right) \geq B_t {c_t}^{1-\gamma} g'(\overline{a}), \tag{40}$$

$$B_t c_t^{1-\gamma}\phi\theta_{t+1} + \frac{d}{d\varepsilon_-}B\left(r_1, ...r_{t-1}, \overline{a} + \eta_t + \varepsilon\right) c_t^{1-\gamma}\times$$

$$\times\sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n+(1-\phi)\overline{a})+\kappa_n-\kappa_{n-1}]}\right] \geq B_t {c_t}^{1-\gamma} g'(\overline{a}),$$

$$\frac{d}{d\varepsilon_-}B\left(r_1, ...r_{t-1}, \overline{a} + \eta_t + \varepsilon\right) \geq \frac{B_t\left(g'(\overline{a}) - \phi\theta_{t+1}\right)}{\sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n+(1-\phi)\overline{a})+\kappa_n-\kappa_{n-1}]}\right]} := \theta_t - \phi\theta_{t+1}.$$

The second equivalence above follows from the fact that for $s > t$

$$E_t\left[c_s^{1-\gamma}\right] = c_t^{1-\gamma} E_t\left[e^{(1-\gamma)\sum_{n=t+1}^{s}[\theta_n(\varepsilon_n+(1-\phi)\overline{a})+\kappa_n-\kappa_{n-1}]}\right] =$$

$$= c_t^{1-\gamma}\prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n+(1-\phi)\overline{a})+\kappa_n-\kappa_{n-1}]}\right].$$

One can inductively show that for any $t \leq L$, $0 \leq \theta_t \leq g'(\overline{a})$. Therefore, proceeding analogously as in the proof for $t = L$, we can establish that indeed (40) holds with equality. Integrating out this equality we establish that for $t' \geq t$,

$$\ln c_{t'}(r_1, ..., r_{t'}) = B(r_1, ..., r_{t-1}) + \theta_{t'}r_{t'} + \sum_{s=t}^{t'-1}(\theta_s - \phi\theta_{s+1})r_s + \kappa_{t'},$$

where θ_s are as required.

We now determine the values of the constants κ_t. First, there exists a value $e^{-\underline{\kappa}}$ such

[14] Equation (39) can also be derived from the IEE if $\gamma = 1$.

that $e^{-\underline{\kappa}} = e^{Rt} B_t E\left[c_t^{-\gamma}\right]$ for $t \leq T$ for all t. This yields, for all t:

$$\gamma \kappa_t = \underline{\kappa} + Rt + \ln B_t + \sum_{s=1}^{t} \ln E\left[e^{-\gamma\theta_s(\varepsilon_s + (1-\phi)\overline{a})}\right].$$

Finally, constant $\underline{\kappa}$ is chosen so that the agent's participation constraint binds, and $\acute{k}_t = \kappa_t - \kappa_{t-1}$. When the PS constraint is not imposed, we use (7) to derive (13) in an analogous way;

Now suppose that the NM constraint is imposed. Proceeding inductively as above we establish that

$$\ln c_t = \sum_{s=1}^{t} \theta_s(r_s - \phi r_{s-1}) + \kappa_t,$$

with $\theta_t = 0$ for $t > L + M$, and k_t as in the Theorem. The θ_t are the lowest values such that the IC and NM constraints are satisfied, i.e.:

$$IC : \theta_t - \phi\theta_{t+1} \geq \frac{B_t\left(g'(\overline{a}) - \phi\theta_{t+1}\right)}{\sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + (1-\phi)\overline{a}) + k_n]}\right]}, \text{ for } 0 \leq t \leq L, \tag{41}$$

$$NM : E_t\left[\frac{\partial U}{\partial r_t}\right] = E_t\left[\frac{\partial U}{\partial r_{t+i}}\right], \text{ for } 0 \leq t \leq L,\ 0 \leq i \leq M. \tag{42}$$

If we set

$$\theta_{L+i} = \frac{D_i}{\sum_{s=L+i}^{T} B_s \prod_{n=L+i+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + (1-\phi)\overline{a}) + k_n]}\right]},$$

for some constants D_i, $i \leq M$, (42) is equivalent to

$$\begin{aligned} & B_t c_t^{1-\gamma}\phi\theta_{t+1} + \theta_t c_t^{1-\gamma} \sum_{s=t}^{T} B_s \prod_{n=t+1}^{s} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + (1-\phi)\overline{a}) + k_n]}\right] = \\ & = E_t\left[c_{L+i}^{1-\gamma}(B_{L+i}\phi\theta_{L+i+1} + D_i)\right] = \\ & = c_t^{1-\gamma} \prod_{n=t+1}^{L+i} E_t\left[e^{(1-\gamma)[\theta_n(\varepsilon_n + (1-\phi)\overline{a}) + k_n]}\right](B_{L+i}\phi\theta_{L+i+1} + D_i), \end{aligned}$$

for $0 \leq t \leq L$, $i \leq M$. This yields the desired expressions for θ_t', $t \leq L + M$, with $D = D_M$.

A.3 Proof of Theorem 3

We divide the proofs into the following steps.*[AE->TS: I have restructured the proof. Please take a look]*

Step 1. Change of variables. Consider the new variable x_t, $t \leq L$, and per period utility functions $u(c_t, x_t)$ defined as:

$$x_t = \begin{cases} -g(a_t) & \text{if } \gamma = 1 \\ e^{-g(a_t)\frac{1-\gamma}{\gamma}}\xi & \text{if } \gamma \neq 1 \end{cases} \quad , \quad u(c_t, x_t) = \begin{cases} \ln c_t + x_t & \text{if } \gamma = 1 \\ \frac{c_t{}^{1-\gamma}(\xi x_t)^{\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \end{cases} \quad ,$$

where $\xi = sign(1-\gamma)$, and let $a_t = f(x_t)$. x_t measures the agent's leisure and f is the "production function" from leisure to effort, which is decreasing and concave. The new variables are chosen in such a way that the CEO's utility is jointly concave in consumption and leisure.

Let $U\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right) = \sum_{t=1}^{T} \rho^t u(c_t, x_t)$ be total discounted utility and consider the maximization problem:

$$\max_{x_t, c_t, m_t \text{ adapted}} E\left[U\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right)\right], \tag{43}$$

with $\sum_{t=1}^{T} e^{-rt}(y_t - c_t) \geq 0$ and income y_t satisfying

$$\ln y_t = \sum_{s=1}^{t} \theta_s(\eta_s + f(x_s) + \overline{m}_s - \phi(\eta_{s-1} + f(x_{s-1}) + \overline{m}_{s-1})) + k_t, \tag{44}$$

for $\overline{m}_s$ defined in (31). Problems (43) and (32) are equivalent: $(x_t)_{t\leq L}$, $(c_t)_{t\leq T}$ and $(m_t)_{t\leq L}$ solve the maximization problem (43) if and only if $(f(x_t))_{t\leq L}$, $(c_t)_{t\leq T}$ and $(m_t)_{t\leq L}$ solve the maximization problem (32). Moreover, the utility function $U\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right)$ is jointly concave in $(c_t)_{t\leq T}$ and $(x_t)_{t\leq L}$.

Step 2. Deriving an "upper linearization" utility function. Consider *[AE: can we call this $\overline{c}$ and get rid of the asterisks, since the $f(x_t^*) = \overline{a}$.XG I think $\overline{c}$evokes maximum consumption, and that's bad. So stars are good]* $c_t^*(\eta) = \exp\left(\sum_{n=1}^{t} \theta_n(\eta_n + f(x_n^*) - \phi(\eta_{s-1} + f(x_{s-1}))) + k_t\right)$, the consumption for the recommended sequence of leisure on the path of noises $\eta = (\eta_t)_{t\leq T}$

(where $a_t^* = f(x_t^*)$), under no saving or manipulation. For any path of noises $\eta = (\eta_t)_{t \leq T}$ we introduce the "upper linearization" utility function $\widehat{U}_\eta$:

$$\widehat{U}_\eta\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right) = U + \sum_{t=1}^{T}\left(c_t - c_t^*(\eta)\right)\frac{\partial U}{\partial c_t} + \sum_{t=1}^{L}(x_t - x_t^*)\frac{\partial U}{\partial x_t}, \tag{45}$$

where U, $\frac{\partial U}{\partial c_t}$ and $\frac{\partial U}{\partial x_t}$ are evaluated at the (noise dependent) target consumption and leisure levels $\left(c_t^*(\eta)\right)_{t\leq T}, (x_t^*)_{t\leq L})$. Since $U = U\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right)$ is jointly concave in $(c_t)_{t\leq T}$ and $(x_t)_{t\leq L}$, we have the key property:

$$\widehat{U}_\eta\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right) \geq U\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right) \text{ for all paths } \eta,\ (c_t)_{t\leq T}, (x_t)_{t\leq L}.$$

$$\widehat{U}_\eta\left((c_t^*(\eta))_{t\leq T}, (x_t^*)_{t\leq L}\right) = U\left((c_t^*(\eta))_{t\leq T}, (x_t^*)_{t\leq L}\right) \text{ for all paths } \eta.$$

Hence, to show that there are no profitable deviations for EU, it is sufficient to show that there are no profitable deviations for $E\widehat{U}_\eta$.

Moreover, since

$$e^{rt}\frac{\partial \widehat{U}_\eta}{\partial c_t} = e^{rt}\frac{\partial U\left((c_t^*(\eta))_{t\leq T}, (x_t^*)_{t\leq L}\right)}{\partial c_t} = \frac{B_t(c_t^*)^{-\gamma}}{e^{-rt}},$$

when private savings are allowed, the PS constraint (6) implies that $e^{rt}\frac{\partial \widehat{U}_\eta}{\partial c_t}$ is a martingale. Therefore, the agent is indifferent at which time he consumes income y_t, and so we can evaluate $E\widehat{U}_\eta$ for $c_t \equiv y_t$. Since it is automatic that the agent has no motive to save, we now only need to show that he has no motive to engage in manipulation or change his choice of leisure (and thus effort). We can also abuse notation and let $\widehat{U}_\eta$ be a function of $(x_t)_{t\leq L}$ and $(m_t)_{t\leq L}$, since they fully determine the process of income $(y_t)_{t\leq T}$ and thus consumption $(c_t)_{t\leq T}$.

The results are summarized in the following Lemma.

Lemma 1 *(Upper linearization.) Let* $\widetilde{U}_\eta\left((m_t)_{t\leq L}, (x_t)_{t\leq L}\right) = \widehat{U}_\eta\left((y_t)_{t\leq T}, (x_t)_{t\leq L}\right)$ *for* $\widehat{U}_\eta$ *defined as in (45) and* y_t *as in (44), and consider the following maximization problem:*

$$\max_{x_t, m_t \text{ adapted}} E\left[\widetilde{U}_\eta\left((m_t)_{t\leq L}, (x_t)_{t\leq L}\right)\right]. \tag{46}$$

If the target leisure level $(x_t^*)_{t\leq L}$ *and no manipulation,* $m_t \equiv \mathbf{0}$, $t \leq L$, *solve the maximization* *problem (46) then* $(c_t^*)_{t\le T}$, $(x_t^*)_{t\le L}$ *and* $m_t \equiv \mathbf{0}$, $t \le L$, *solve the maximization problem (43).*

Step 3. Pathwise concavity of utility in leisure and manipulation for $\gamma = 1$. To show that agent has no incentive to engage in manipulation or change his effort choice, we must demonstrate that expected utility is jointly concave in leisure $(x_t)_{t\le L}$ and manipulations $(m_t)_{t\le L}$. For $\gamma = 1$, we can do so by proving pathwise concavity of utility in leisure and manipulation. (We will deal with the case $\gamma \neq 1$ in step 4). For any $\gamma > 0$ and every path η the linearized utility function with no savings $\widetilde{U}_\eta$ has the form:

$$\widetilde{U}_\eta\left((m_t)_{t\le L},(x_t)_{t\le L}\right) = A+\sum_{t=1}^{L} B_t x_t+\sum_{t=1}^{T} C_t(\eta) e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1}))+t\ln\rho}, \tag{47}$$

for some constants A, B_t and random variables $C_t(\eta)$. For example, if $\gamma = 1$, we have:*[AE -> TS: I think the first below should be a sum to L, and the* $\rho^t x_t$ *term should be in brackets]*

$$\widetilde{U}_\eta\left((m_t)_{t\le L},(x_t)_{t\le L}\right) = \sum_{t=1}^{T}\rho^t(\ln c_t^*(\eta)-1)+\rho^t x_t+\sum_{t=1}^{T} e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1}))+t\ln\rho}. \tag{48}$$

To prove that (48) is jointly concave in $(x_t)_{t\le L}$ and $(m_t)_{t\le L}$, we must show that the "PV of income function"

$$I\left((m_t)_{t\le L},(x_t)_{t\le L}\right) = \sum_{t=1}^{T} e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1}))+t\ln\rho}$$

is concave. For this we will use the following general Lemma, proven at the end of this section. *[AE->TS: I have redefined variables here from x to b. b is a M+1 vector which contains both the scalar x and the M-vector m. Previously, there was overloading of the term x which was being used for both an M+1 vector in the lemma and a scalar outside]*

Lemma 2 *(Concavity of present values.) Let*

$$I((b_t)_{t\le T}) = \sum_{t=1}^{T}\exp\left(\sum_{s=1}^{t-M} j_s(b_s)+\sum_{s=t-M+1}^{t} q_s^t(b_s)\right),$$

where $b_s \in \mathbb{R}^{M+1}$ *and all* j_s *and* q_s^t *are twice differentiable functions with* $\frac{\partial}{\partial \mathbf{b}_{s,i}\partial \mathbf{b}_{s,k}} j_s =$

$\frac{\partial}{\partial \mathbf{b}_{s,i}\partial \mathbf{b}_{s,k}} q_s^t = 0$, $\frac{\partial}{\partial \mathbf{b}_{s,i}} j_s \leq \frac{\partial}{\partial \mathbf{b}_{s,i}} q_s^t$. *Suppose that for every s:*

$$\sup \left[2(M+C)(M+1)^2 \left(\frac{\partial}{\partial \mathbf{b}_{s,i}} q_s^n \right)^2 + \frac{\partial^2}{(\partial \mathbf{b}_{s,i})^2} q_s^n \right] \leq 0, \; i \leq M+1, \; n \leq s+M \tag{49}$$
$$\sup \left[2C(M+1)^2 \left(\frac{\partial}{\partial \mathbf{b}_{s,i}} j_s \right)^2 + \frac{\partial^2}{(\partial \mathbf{b}_{s,i})^2} j_s \right] \leq 0, \; i \leq M+1,$$

for $C = e^{M(\sup q_s^t - \inf q_s^t)/2} \sum_{n=0}^{T} e^{n \sup j_t /2}$, *and at least one of these inequalities is strict. Then the function* I *is concave.*

Loosely speaking, the Lemma states that, if j_s and q_t are sufficiently concave functions, then the "present value of income" function $I\left((b_t)_{t\leq L}\right)$ associated with them is also jointly concave in the sequence of decisions $(b_t)_{t\leq L}$. (The decision vector b is an $M+1$-vector that incorporates both the scalar x and the M-vector m.) This is non-trivial to prove when T is infinite: for sufficiently large t, $\exp\left(tj\left(b\right)\right)$ is a convex function of b, because its second derivative (when b is one-dimensional) is $\exp\left(tj\left(b\right)\right) t \left(tj'\left(b\right)^2 + j''\left(b\right)\right)$, which is positive for sufficiently large t. It is discounting (expressed by $\delta < 1$) that allows the income function to be concave.

To show that $I\left((m_t)_{t\leq L}, (x_t)_{t\leq L}\right)$ is jointly concave in leisure $(x_t)_{t\leq L}$ and manipulations $(m_t)_{t\leq L}$ we use Lemma 2 with $b_t = (m_t, x_t)$ and:

$$j_s(x_s, m_s) = (\theta_s - \phi\theta_{s+1}) \left[f(x_s) - \overline{a} + \sum_{i=1}^{M} (m_{s,i} - \lambda(m_{s,i})) \right] - \sum_{i=1}^{M} (\theta_{s+i} - \phi\theta_{s+i+1}) m_{s,i} + \ln\rho, \tag{50}$$
$$q_s^t(x_s, m_s) = (\theta_s - \phi\theta_{s+1}) \left[f(x_s) - \overline{a} + \sum_{i=1}^{M} (m_{s,i} - \lambda(m_{s,i})) \right] - \sum_{i=1}^{t-s} (\theta_{s+i} - \phi\theta_{s+i+1}) m_{s,i} + \ln\rho, \;\; s < t$$
$$q_t^t(x_t, m_t) = \theta_s \left[f(x_s) - \overline{a} + \sum_{i=1}^{M} (m_{s,i} - \lambda(m_{s,i})) \right] + \ln\rho.$$

Step 4. Concavity of expected utility in leisure and manipulation for $\gamma \neq 1$. When $\gamma \neq 1$ then the linearized utility $\widehat{U}_\eta$ is:

$$\widehat{U}_\eta\left((c_t)_{t\le T},(x_t)_{t\le L}\right)=\sum_{t=1}^{L}\frac{\gamma}{1-\gamma}\rho^t c_t^*(\eta)^{1-\gamma}\left(\frac{x_t}{(\xi x_t^*)^{1-\gamma}}\right)+\sum_{t=1}^{T}\rho^t(\xi x_t^*)^\gamma\left(\frac{c_t}{c_t^{*\gamma}(\eta)}\right), \quad (51a)$$

with $\xi = sign(1-\gamma)$. Unlike equation (48) in the case of $\gamma = 1$, linearized utility $\widehat{U}_\eta$ now depends on noise η. We therefore are unable to prove pathwise concavity of linearized utility, and instead prove concavity of expected utility directly. Linearized utility with no savings $\widetilde{U}_\eta$ has the form (47) with

$$C_t(\eta)=e^{(1-\gamma)\left[k_t-g(a_t^*)+\sum_{n=1}^{t}(\theta_n-\phi\theta_{n+1})\varepsilon_n\right]}=M_t(\eta)C_t', \quad (52)$$

where $M_t(\eta)=e^{\sum_{n=1}^{t}\left[(1-\gamma)(\theta_n-\phi\theta_{n+1})\varepsilon_n-\ln E\left(e^{(1-\gamma)(\theta_n-\phi\theta_{n+1})\varepsilon_n}\right)\right]}$ is a martingale and

$$C_t'=e^{(1-\gamma)[k_t-g(a_t^*)]+\sum_{n=1}^{t}\ln E\left(e^{(1-\gamma)(\theta_n-\phi\theta_{n+1})\varepsilon_n}\right)}.$$

Expected utility is given by

$$E\left[\widetilde{U}_\eta((m_t)_{t\le L},(x_t)_{t\le L})\right]=E\left[A+\sum_{t=1}^{L}B_t x_t+\sum_{t=1}^{T}M_t(\eta)C_t' e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1}))+t\ln\rho}\right]$$
$$=E\left[A+\sum_{t=1}^{L}B_t x_t+M_T(\eta)\sum_{t=1}^{T}C_t' e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1}))+t\ln\rho}\right.$$

where the second equality follows from the law of iterated expectations and the fact that $M_T(\eta)$ is a martingale. Two paragraphs below we use Lemma 2 to show that the modified "present value of income" function $I'((m_t)_{t\le L},(x_t)_{t\le L})=\sum_{t=1}^{T}C_t' e^{\sum_{n=1}^{t}\theta_n(f(x_n)-\gamma a_n^*+\overline{m}_n-\phi(f(x_{n-1})-\gamma a_{n-1}^*+\overline{m}_{n-1})}$ for $\overline{m}_s$ defined in (31), is pathwise jointly concave in leisure and manipulation. Therefore, $E\widetilde{U}_\eta$ is concave in the processes $(x_t)_{t\le L}$ and $(m_t)_{t\le L}$.

We now conclude the proof of the Theorem. From Theorem 2, $E\widetilde{U}_\eta$ satisfies the first-order conditions at $(x_t^*)_{t\le L}$ and $(m_t)_{t\le L}$. From step 4, $E\widetilde{U}_\eta$ is also concave in $(x_t)_{t\le L}$ and $(m_t)_{t\le L}$, and so the target leisure level $(x_t^*)_{t\le L}$ and no manipulations, $m_t\equiv\mathbf{0}$, $t\le L$, solve the maximization problem (46). Therefore, from Lemma 1, $(c_t^*)_{t\le T}$, $(x_t^*)_{t\le L}$ and $m_t\equiv\mathbf{0}$, $t\le L$, solve the maximization problem (43), establishing the result.

[AE->TS: should we move this up to after "the fact that $M_T(\eta)$ is a martingale."? Here it is a little out of order; this proof is not that long, so I think we can put it in its correct place] It remains to show that both I and I' are concave in $(x_t)_{t\le L}$ and $(m_t)_{t\le L}$. In the first case we use Lemma 2 for j_s and q_s^t defined as in (50). We have $\frac{\theta_t-\phi\theta_{t+1}}{\theta_s-\phi\theta_{s+1}} \le D_0$ as long as $|t-s|\le M$, for some $D_0>0$. Let λ be such that $\sup D_0 m-\lambda(m)\le D_1$, for some $D_1>0$, and m^* be such that $D_0 m-\lambda(m)\le 0$ for $m\ge m^*$. We can assume without loss of generality that the CEO chooses manipulations only within the interval $[-m^*,m^*]$, and so $C=e^{M(\sup q_s^t-\inf q_s^t)/2}\sum_{n=0}^T e^{n\sup j_t/2}$ is finite. Finally, since $f'(x_s)=\frac{-1}{g'(f(x_s))}$, $f''(x_s)=\frac{-g''(f(x_s))}{g'^3(f(x_s))}$ and $\theta_s\le g'(\overline{a})$, the condition (49) is satisfied for $i=1$ if g has sufficiently high curvature. Moreover, since $\frac{\partial}{\partial m_{s,i}}q_s^n=(\theta_s-\phi\theta_{s+1})(1-\lambda'(m_{s,i}))-\mathbf{1}_{n<i}(\theta_{s+i}-\phi\theta_{s+i+1})$, $\frac{\partial}{\partial m_{s,i}}j_s=(\theta_s-\phi\theta_{s+1})(1-\lambda'(m_{s,i})$ and $\frac{\partial^2}{(\partial m_{s,i})^2}q_s^n=\frac{\partial^2}{(\partial m_{s,i})^2}j_s=-(\theta_s-\phi\theta_{s+1})\lambda''(m_{s,i})$, the condition (49) is satisfied for $i>1$ if λ has sufficiently high curvature.

In the case of I' we must verify condition (49) in Lemma 2 when j_s and q_s^t are defined as:

$$j_s(x_s,m_s)=\theta_s\left[f(x_s)-\gamma a_t^*+\sum_{i=1}^M(m_{s,i}-\lambda(m_{s,i}))\right]-\sum_{i=1}^M\theta_{s+i}m_{s,i}+D_s, \tag{53}$$

$$q_s^t(x_s,m_s)=\theta_s\left[f(x_s)-\gamma a_t^*+\sum_{i=1}^M(m_{s,i}-\lambda(m_{s,i}))\right]-\sum_{i=1}^{t-s}\theta_{s+i}m_{s,i}+D_s,\ s<t,$$

for $D_s=(1-\gamma)\left[(k_s-g(a_s^*))-(k_{s-1}-g(a_{s-1}^*))\right]+\ln E\left(e^{(1-\gamma)\theta_n e_n}\right)+\ln\rho$. The rest of the proof follows just as in the $\gamma=1$ case, with the derivatives of the f function being:

$$f'(x_s)=-D\frac{1}{x_s g'(f(x_s))},\ f''(x_s)=\frac{1}{x_s^2 g'^2(f(x_s))}\left(Dg'(f(x_s))-D^2\frac{g''(f(x_s))}{g'(f(x_s))}\right),$$

for $D=\frac{\gamma}{1-\gamma}sign(1-\gamma)$.

Proof of Lemma 2 Let

$$P_s((x_t)_{t\le T})=e^{\sum_{n=1}^{s-M}j_n(x_n)+\sum_{n=s-M+1}^{s}q_n^s(x_n)},$$

$$S_s((x_t)_{t\le T})=\sum_{n=s}^T e^{\sum_{m=1}^{n-M}j_m(x_m)+\sum_{m=n-M+1}^{n}q_m^n(x_m)}=\sum_{n=s}^T P_n((x_t)_{t\le L}),$$

for any $s\le T$. For the rest of the proof, fix an argument sequence $(x_t)_{t\le T}$. We will evaluate all the functions at this sequence, and consequently economize on notation by dropping the argument of S_s and $q_{t,s}$.

Step 1: Derivatives. For unit vectors e_r^i and e_s^k, $r \geq s$, $i,k \leq M+1$, consider the derivatives of the function I:

$$\frac{\partial I}{\partial e_s^k} = \sum_{n=s}^{s+M-1} \partial_k q_s^n P_n + \partial_k j_s S_{s+M},$$

$$\frac{\partial^2 I}{\partial e_r^i \partial e_s^k} = \sum_{n=r}^{s+M-1} \partial_k q_s^n \partial_i q_r^n P_n + \partial_k j_s \left(\sum_{n=\max\{r,s+M\}}^{r+M-1} \partial_i q_r^n P_n + \partial_i j_r S_{r+M} \right) + $$
$$+ \mathbf{1}_{r=s,i=k} \left[\sum_{n=s}^{s+M-1} \partial_k^2 q_s^n P_n + \partial_k^2 j_s S_{s+M} \right],$$

where we define $\partial_k f(x) = \frac{\partial}{\partial x_k} f(x)$ and $\partial_k^2 f(x) = \frac{\partial^2}{(\partial x_k)} f(x)$. Therefore, for a fixed vector $y = (y_t)_{t \leq T}$ the second derivative in the direction $y = (y_t)_{t \leq T}$ is:

$$\frac{\partial^2 I}{\partial y \partial y} = \sum_{k,i=1}^{M+1} \sum_{s=1}^{T} \sum_{r=1}^{T} y_s^k y_r^i \frac{\partial^2 I}{\partial e_s^k \partial e_r^i} =$$
$$= 2 \sum_{k,i=1}^{M+1} \sum_{s=1}^{T} \sum_{r \geq s} y_s^k y_r^i \left[\sum_{n=r}^{s+M-1} \partial_k q_s^n \partial_i q_r^n P_n + \partial_k j_s \left(\sum_{n=\max\{r,s+M\}}^{r+M-1} \partial_i q_r^n P_n + \partial_i j_r S_{r+M} \right) \right]$$
$$+ \sum_{i=1}^{M+1} \sum_{s=1}^{T} y_s^{i2} \left[\sum_{n=s}^{s+M-1} \partial_i^2 q_s^n P_n + \partial_i^2 j_s S_{s+M} \right] =: W + V.$$

Step 2: Bounding P_r and S_r. For any $s \leq T$ and $q \leq T - s$ we have:

$$P_{s+q} = e^{\sum_{n=1}^{s+q} j_n + \sum_{n=s+q-M+1}^{s} q_n^s} \leq e^{M \sup q_t} e^{\sum_{n=1}^{s+q} j_n} \leq e^{M \sup q_t + q \sup j_t} e^{\sum_{n=1}^{s} j_n} \leq e^{q \sup j_t + M(\sup q_t - \inf q_t)} P_s,$$

It follows that for $\psi = \frac{\sup j_s}{2}$ we have:

$$\sum_{r \geq s} P_r e^{-\psi(r-s)} \leq C_1 P_s, \quad \sum_{s,r \geq s} P_r y_r^2 e^{\psi(r-s)} = \sum_r y_r^2 P_r \sum_{s \leq r} e^{\psi(r-s)} \leq C_2 \sum_s P_s y_s^2, \tag{54}$$

where

$$C_1 = e^{M(\sup q_t - \inf q_t)} \sum_{n=0}^{T} e^{n\psi}, \; C_2 = \sum_{n=0}^{T} e^{n\psi}. \tag{55}$$

Moreover, since $S_{s+q} \leq e^{q \sup j_t + M(\sup q_t - \inf q_t)} S_s$, the inequalities for S_r analogous to (54) also hold.

Step 3: Bounding the derivatives. For any vector $z{=}(z_t)_{t \leq T}$, $z_t \in \mathbb{R}$, we have:

$$\sum_{s,r\geq s} z_s z_r P_r = \sum_s z_s \sum_{r\geq s} \sqrt{P_r} z_r e^{\frac{\psi}{2}(r-s)} \sqrt{P_r} e^{-\frac{\psi}{2}(r-s)} \leq \sum_s z_s \left(\sum_{r\geq s} P_r z_r^2 e^{\psi(r-s)}\right)^{1/2} \left(\sum_{r\geq s} P_r e^{-\psi(r-s)}\right)^{1/2} \leq$$
$$\leq \sqrt{C_1} \sum_s z_s \sqrt{P_s} \left(\sum_{r\geq s} P_r z_r^2 e^{\psi(r-s)}\right)^{1/2} \leq \sqrt{C_1} \left(\sum_s z_s^2 P_s\right)^{1/2} \left(\sum_s \left(\sum_{r\geq s} P_r z_r^2 e^{\psi(r-s)}\right)\right)^{1/2} \leq$$
$$\leq \sqrt{C_1 C_2} \left(\sum_s z_s^2 P_s\right)^{1/2} \left(\sum_s P_s z_s^2\right)^{1/2} = C \sum_s z_s^2 P_s,$$

where the first and third inequalities follow from the Cauchy-Schwartz inequality, and C_1 and C_2 are as in (55). Similarly, we obtain $\sum_{s,r\geq s} z_s z_r S_r \leq C \sum_s z_s^2 S_s$. Therefore:

$$W = 2 \sum_{k,i=1}^{M+1} \sum_{s=1}^{T} \sum_{r\geq s} y_s^k y_r^i \left[\sum_{n=r}^{s+M-1} \partial_k q_s^n \partial_i q_r^n P_n + \partial_k j_s \left(\sum_{n=\max\{r,s+M\}}^{r+M-1} \partial_i q_r^n P_n + \partial_i j_r S_{r+M} \right) \right]$$
$$\leq 2 \sum_{k,i=1}^{M+1} \left\{ \sum_{n=1}^{T} P_n \left[\sum_{s\geq n-M, r\geq s} y_s^k \partial_k q_s^n y_r^i \partial_i q_r^n \right] + \sum_{m=0}^{M-1} \sum_{s=1}^{T} \sum_{r\geq s} \left[y_s^k \partial_k q_s^{s+m} y_r^i \partial_i q_r^{r+m} P_{r+m} \right] + \right.$$
$$\left. + \sum_{s=1}^{T} \sum_{r\geq s} \left[y_s^k \partial_k j_s y_r^i \partial_i j_r S_{r+M} \right] \right\}$$
$$\leq 2(M+1)^2 \left\{ \sum_{n=1}^{T} P_n \left[\sum_{s\geq n-M, r\geq s} \max_i (y_s^i \partial_i q_s^n) \max_i (y_r^i \partial_i q_r^n) \right] \right.$$
$$\left. + \sum_{m=0}^{M-1} \sum_{s=1}^{T} \sum_{r\geq s} \left[\max_i (y_s^i \partial_i q_s^{s+m}) \max_i (y_r^i \partial_i q_r^{r+m}) P_{r+m} \right] + \sum_{s=1}^{T} \sum_{r\geq s} \left[\max_i (y_s^i \partial_i j_s) \max_i (y_r^i \partial_i j_r) S_{r+M} \right] \right\}$$
$$\leq 2(M+1)^2 \left\{ \sum_{n=1}^{T} P_n \left[\sum_{s\geq n-M} M \max_i (y_s^i \partial_i q_s^n)^2 \right] + \sum_{m=0}^{M-1} \sum_{s=1}^{T} \left[C \max_i (y_s^i \partial_i q_s^{s+m})^2 P_{s+m} \right] \right.$$
$$\left. + \sum_{s=1}^{T} \left[C \max_i (y_s^i \partial_i j_s)^2 S_{r+M} \right] \right\}$$
$$\leq 2(M+1)^2 \sum_{s=1}^{T} \sum_{i=1}^{M+1} y_s^{i2} \left[\sum_{m=0}^{M-1} (M+C)(\partial_i q_s^{s+m})^2 P_{s+m} + C(\partial_i j_s)^2 S_{r+M} \right].$$

Finally,

$$\frac{\partial^2 I}{\partial y \partial y} = W + V$$
$$\leq \sum_{s=1}^{T} \sum_{i=1}^{M+1} y_s^{i2} \left[\sum_{n=s}^{s+M-1} \left(2(M+C)(M+1)^2 \left(\partial_i q_s^n\right)^2 + \partial_i^2 q_s^n\right) P_n + \left(2C(M+1)^2 \left(\partial_i j_s\right)^2 + \partial_i^2 j_s\right) S_s \right],$$

establishing the Lemma. ■

B Variable Cost of Effort

This section extends the core model to allow a deterministically varying marginal cost of effort. In practice, this occurs if either the cost function or maximum effort level changes over time. For example, for a start-up firm, the CEO can undertake many actions to improve firm value (augmenting the maximum effort level) and effort is relatively productive (reducing the cost of effort).

We now allow for a time-varying maximum effort level $\bar{a}_t$ and cost of effort $g_t(\cdot)$. The slope of the contract in Theorem 1 (equations (11) and (12)) now becomes:

$$\theta_t = \begin{cases} \frac{g_t'(\bar{a}_t)}{1+\rho+\ldots\rho^{T-t}} \text{ for } t \leq L, \\ 0 \text{ for } t > L \end{cases} \tag{56}$$

if manipulation is impossible, and if manipulation is possible

$$\theta_t = \begin{cases} \theta_t = \frac{\Theta}{1+\rho+\ldots\rho^{T-t}} \rho^{-t} \text{ for } t \leq L+M, \\ 0 \text{ for } t > L+M \end{cases} . \tag{57}$$

where $\Theta = \sup_{s \leq L} \left(\rho^s g_s'\left(\bar{a}_s\right)\right)$.

We previously showed that imposing the NM constraint causes the contract's slope to rise over time; the speed of the rise depended only on the CEO's impatience ρ. With a non-constant target action, it depends on $\Theta = \sup_{s \leq L} \left(\rho^s g_s'\left(\bar{a}_s\right)\right)$, the maximum discounted sensitivity during the CEO's working life. Let $s \leq L$ denote the period in which $\rho^s g_s'\left(\bar{a}_s\right)$ is highest. The CEO has an incentive to increase r_s at the expense of the signal in any t within M periods of s. Therefore, the sensitivity for all t within M periods of s must increase, to remove these incentives. However, this in turn has a knock-on effect: since the sensitivity for $t = s - M$ has now risen, the CEO now has an incentive to increase r_{s-M} at the expense of r_{s-2M}, and so on. Therefore, the sensitivity at s forces upward the sensitivity in all periods $t \leq L + M$, even those more than M periods away from s, because of the knock-on effects. This explains why the contract in all periods $t \leq L + M$ depends on Θ in equation (57).

This dependence can be illustrated in a numerical example. We first set $T = 5$, $L = 3$, $\rho = 1$, $g_1'(\overline{a}_1) = g_2'(\overline{a}_2) = 1$ and $g_3'(\overline{a}_3) = 2$. If manipulation is impossible, the optimal contract is

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3} r_3 + \kappa_3 \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3} r_3 + \kappa_4 \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3} r_3 + \kappa_5 .
\end{aligned}$$

Since the marginal cost of effort is high at $t = 3$, the contract sensitivity must be high at $t = 3$ to satisfy the IC condition. However, this now gives the CEO incentives to engage in manipulation if it were possible. If he manipulates r_2 downwards by 1 unit to augment r_3 by 1 unit, lifetime consumption falls by 1 unit and rises by 2 units. Therefore, the sensitivity of the contract at $t = 2$ must increase to remove these incentives. This increased sensitivity at $t = 2$ in turn augments the required sensitivity at $t = 1$, else the CEO would manipulate to reduce r_1 and increase r_2. Therefore, even though the maximum release lag M is 1 and so the CEO cannot directly manipulate r_1 to affect r_3, the high sensitivity at r_3 still affects the sensitivity at r_1 by changing the sensitivity at r_2. The new contract is given by:

$$\ln c_1 = \frac{2}{5}r_1 + \kappa_1$$
$$\ln c_2 = \frac{2}{5}r_1 + \frac{r_2}{2} + \kappa_2$$
$$\ln c_3 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + \kappa_3$$
$$\ln c_4 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + r_4 + \kappa_4$$
$$\ln c_5 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + r_4 + \kappa_5.$$

C Continuous Time

We now consider the continuous-time analog of the model. The CEO's utility is given by:

$$U = \begin{cases} E\left[\int_0^T \rho^t \frac{(c_t h(a_t))^{1-\gamma}-1}{1-\gamma} dt\right] & \text{if } \gamma \neq 1 \\ E\left[\int_0^T \rho^t \left(\ln c_t + \ln h\left(a_t\right)\right) dt\right] & \text{if } \gamma = 1. \end{cases} \tag{58}$$

For now, we consider the log utility case; in a later draft we will extend this section to general CRRA utility functions. The firm's returns evolve according to:

$$dR_t = a_t dt + \sigma_t dZ_t$$

where Z_t is a Brownian motion, and the volatility process σ_t is deterministic. We normalize $r_0 = 0$ and the risk premium to zero, i.e. the expected rate of return on the stock is R in each period.

Proposition 1 *(Optimal contract, continuous time, log utility). Let σ_t denote the stock volatility. The optimal contract pays the CEO c_t at each instant, where c_t satisfies:*

$$\ln c_t = \int_0^t \theta_s dR_s + k_{t,} \tag{59}$$

where θ_s and k_t are deterministic functions. If manipulation is impossible, the slope θ_t is *given by:*

$$\theta_t = \begin{cases} \frac{g'(\bar{a})}{\int_t^T \rho^{\tau-s} ds} & \text{for } t \leq L \\ 0 & \text{for } t > L. \end{cases} \tag{60}$$

If manipulation is possible, θ_t is given by:

$$\theta_t = \begin{cases} \frac{g'(\bar{a})\rho^{-t}}{\int_t^T \rho^{\tau-s} ds} & \text{for } t \leq L + M \\ 0 & \text{for } t > L + M. \end{cases} \tag{61}$$

Let $\zeta = -1$ denote the case if private savings are ruled out (and so the PS constraint is not imposed), and $\zeta = 1$ if they are allowed (and so the PS constraint is imposed). The value of k_t is:

$$k_t = (R + \ln \rho)\, t - \int_0^t \theta_s E\left[dR_s\right] + \zeta \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds + \underline{k}, \tag{62}$$

where $\underline{k}$ ensures that the agent is at his reservation utility.

The implications of the optimal contract are the same as for discrete time, except that the rebalancing of the account is now continuous.

Proposition 2 *(Optimal contract, continuous time, general CRRA utility, with Private Savings constraint). Let σ_t denote the stock volatility. The optimal contract pays the CEO c_t at each instant, where c_t satisfies:*

$$\ln c_t = \int_0^t \theta_s dR_s + \kappa_t, \tag{63}$$

where θ_s and κ_t are deterministic functions. If manipulation is impossible, the slope θ_t is given by:

$$\theta_t = \frac{\rho^t e^{-(1-\gamma)g(\bar{a})} g'(\bar{a})}{\int_t^T \rho^s e^{-(1-\gamma)g(\bar{a})+(1-\gamma)(k_s-k_t)} E_t\left[e^{(1-\gamma)\int_t^s \theta_\tau dR_\tau}\right] ds} \quad \text{for } t \leq L, \tag{64}$$

$$\theta_t = 0 \quad \text{for } t > L.$$

If manipulation is possible, θ_t is given by:

$$\theta_t = \frac{De^{(1-\gamma)(k_{L+M}-k_t)} E_t\left[e^{(1-\gamma)\int_t^{L+M}\theta_\tau dR_\tau}\right]}{\int_t^T \rho^s e^{-(1-\gamma)g(\bar{a})+(1-\gamma)(k_s-k_t)} E_t\left[e^{(1-\gamma)\int_t^s \theta_\tau dR_\tau}\right] ds} \qquad \textit{for } t \leq L+M,$$

$$\theta_t = 0 \qquad \textit{for } t > L+M.$$

The value of κ_t is:

$$\gamma\kappa_t = (r + \ln\rho)t - (1-\gamma)g(\overline{a}) - \gamma\int_0^t \theta_s \overline{a} ds + \frac{1}{2}\gamma^2 \int_0^t \theta_s^2 \sigma_s^2 ds + \underline{\kappa}, \tag{65}$$

where $\underline{\kappa}$ ensures that the agent is at his reservation utility, and D is the lowest constant such that:

$$De^{(1-\gamma)(k_{L+M}-k_t)} E_t\left[e^{(1-\gamma)\int_t^{L+M}\theta_\tau dR_\tau}\right] \geq \rho^t e^{-(1-\gamma)g(\bar{a})} g'(\bar{a}), \ \textit{for all } t \leq L.$$

References

[1] Atkeson, Andrew and Robert E. Lucas Jr (1992): "On Efficient Distribution with Private Information," *Review of Economic Studies* 59, 427-53.

[2] .Baker, George (1992): "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100, 598-614

[3] Bebchuk, Lucian and Jesse Fried (2004): "Pay Without Performance: The Unfulfilled Promise of Executive Compensation." Harvard University Press, Cambridge.

[4] Bettis, Carr, John Bizjak, Jeffrey Coles and Swaminathan Kalpathy (2008): "Stock and Option Grants with Performance-Based Vesting Provisions." Working Paper, Arizona State University

[5] Bhattacharyya, Sugato and Jonathan Cohn (2008): "The Temporal Structure of Equity Compensation." Working Paper, University of Michigan

[6] Bolton, Patrick, Jose Scheinkman and Wei Xiong (2006): "Executive Compensation and Short-Termist Behaviour in Speculative Markets." *Review of Economic Studies* 73, 577-610

[7] Boschen, John and Kimberly Smith (1995): "You Can Pay Me Now and You Can Pay Me Later: The Dynamic Response of Executive Compensation to Firm Performance." *Journal of Business* 68, 577-608

[8] Caplin, Andrew and Barry Nalebuff (1991): "Aggregation and Social Choice: A Mean Voter Theorem." *Econometrica* 59, 1-23

[9] DeMarzo, Peter and Michael Fishman (2007): "Agency and Optimal Investment Dynamics." *Review of Financial Studies* 20, 151-188.

[10] DeMarzo, Peter and Yuliy Sannikov (2006): "Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model." *Journal of Finance* 61, 2681-2724.

[11] Dittmann, Ingolf, and Ernst Maug (2007): "Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 62, 303-343

[12] Dittmann, Ingolf, Ernst Maug and Oliver Spalt (2008): "Sticks or Carrots? Optimal CEO Compensation when Managers are Loss-Averse." Working Paper, Erasmus University Rotterdam

[13] Edmans, Alex (2009): "Blockholder Trading, Market Efficiency, and Managerial Myopia." *Journal of Finance*, forthcoming.

[14] Edmans, Alex and Xavier Gabaix (2008): "Tractability and Detail-Neutrality in Incentive Contracting." Working Paper, New York University and University of Pennsylvania.

[15] Edmans, Alex, Xavier Gabaix and Augustin Landier (2009): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies*, forthcoming.

[16] Fahlenbrach, Rüdiger and René Stulz (2008): "Managerial Ownership Dynamics and Firm Value." *Journal of Financial Economics*, forthcoming.

[17] Farhi, Emmanuel and Ivan Werning (2009): "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation." Working Paper, Harvard University.

[18] Fong, Kyna and Yuliy Sannikov (2009): "Incentives in Organizational Hierarchies." Working paper, University of California at Berkeley.

[19] Gabaix, Xavier and Augustin Landier (2009): "Why Has CEO Pay Increased So Much?." *Quarterly Journal of Economics* 123, 49-100.

[20] Garrett, Daniel and Alessandro Pavan (2009): "Dynamic Managerial Compensation: A Mechanism Design Approach." Working Paper, Northwestern University.

[21] Gibbons, Robert and Kevin J. Murphy (1992): "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100, 468-505.

[22] Goldman, Eitan and Steve Slezak (2006): "An Equilibrium Model of Incentive Contracts in the Presence of Information Manipulation." *Journal of Financial Economics* 80, 603-626.

[23] Golosov, Mikhail, Narayana Kocherlakota and Aleh Tsyvinski (2003): "Optimal Indirect Capital Taxation." *Review of Economic Studies*, 70, 569–587.

[24] Hall, Brian and Jeffrey Liebman (1998): "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics*, 113, 653-691.

[25] Harris, Milton and Artur Raviv (1979): "Optimal Incentive Contracts With Imperfect Information." *Journal of Economic Theory* 20, 231-259.

[26] He, Zhiguo (2008a): "Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion." *Review of Financial Studies*, forthcoming.

[27] He, Zhiguo (2008b): "Dynamic Compensation Contracts With Private Savings." Working Paper, University of Chicago.

[28] Holmstrom, Bengt (2005): "Pay Without Performance and the Managerial Power Hypothesis: A Comment." *Journal of Corporation Law* 30, 703-715.

[29] Jensen, Michael and Kevin J. Murphy (1990): "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98, 225-64.

[30] Johnson, Shane, Harley Ryan and Yisong Tian (2009): "Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter." *Review of Finance* 13, 115-145.

[31] Kole, Stacey (1997): "The Complexity of Compensation Contracts." *Journal of Financial Economics* 43, 79-104.

[32] Lacker, Jeffrey and John Weinberg (1989): "Optimal Contracts under Costly State Falsification." *Journal of Political Economy* 97, 1345-1363.

[33] Laffont, Jean-Jacques and Jean Tirole (1986): "Using Cost Observation to Regulate Firms." *Journal of Political Economy* 94, 614-641

[34] Murphy, Kevin (1999): "Executive Compensation" in Orley Ashenfelter and David Card, eds., *Handbook of Labor Economics*, Vol. 3b. New York and Oxford: Elsevier/North-Holland, 2485-2563.

[35] Peng, Lin and Ailsa Roell (2008): "Manipulation and Equity-Based Compensation." *American Economic Review P&P* 98, 285-290.

[36] Peng, Lin and Ailsa Roell (2009): "Managerial Incentives and Stock Price Manipulation." Working Paper, Baruch College.

[37] Phelan, Christopher and Andrzej Skrzypacz (2008): "Sequential Equilibria with Infinite Histories," Working Paper, Federal Reserve Bank of Minneapolis.

[38] Rogerson, William P. (1985): "Repeated Moral Hazard." *Econometrica*, 53, 69–76.

[39] Sannikov, Yuliy (2008): "A Continuous-Time Version of the Principal-Agent Problem." *Review of Economic Studies*, 75, 957-984.

[40] Sappington (1983): "Limited Liability Contracts Between Principal and Agent." *Journal of Economic Theory* 29, 1-21.

[41] Shimer, Robert and Ivan Werning (2008): "Liquidity and Insurance for the Unemployed." *American Economic Review* 98, 1922-42.

[42] Stein, Jeremy (1988): "Takeover Threats and Managerial Myopia." *Journal of Political Economy* 46, 61-80.

[43] Werning, Ivan (2001): "Repeated Moral Hazard with Unmonitored Wealth: A Recursive First-Order Approach." Working Paper, Massachusetts Institute of Technology.

[44] Williams, Noah (2008): "On Dynamic Principal-Agent Problems in Continuous Time." Working Paper, University of Wisconsin-Madison.

The Rodney L. White Center for Financial Research
The Wharton School
University of Pennsylvania
3254 Steinberg Hall-Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104-6367

(215) 898-7616
(215) 573-8084 Fax
http://finance.wharton.upenn.edu/~rlwctr

The Rodney L. White Center for Financial Research is one of the oldest financial research centers in the country. It was founded in 1969 through a grant from Oppenheimer & Company in honor of its late partner, Rodney L. White. The Center receives support from its endowment and from annual contributions from its Members.

The Center sponsors a wide range of financial research. It publishes a working paper series and a reprint series. It holds an annual seminar, which for the last several years has focused on household financial decision making.

The Members of the Center gain the opportunity to participate in innovative research to break new ground in the field of finance. Through their membership, they also gain access to the Wharton School's faculty and enjoy other special benefits.

Members of the Center

2009 – 2010

Directing Members

Aronson + Johnson + Ortiz, LP
Ballyshannon Partners
The Nasdaq OMX Educational Foundation
Ther Terker Family
The Vanguard Group

Founding Members

Ford Motor Company Fund
Merrill Lynch, Pierce, Fenner & Smith, Inc.
Oppenheimer & Company
Philadelphia National Bank
Salomon Brothers
Weiss, Peck and Greer