Dynamic CEO Compensation*

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

March 4, 2011

Abstract

We study optimal executive compensation in a fully dynamic framework where the CEO consumes in multiple periods, can undo the contract by privately saving, and can also temporarily inflate the stock price. Despite the complex setup, we obtain a simple closed-form contract. It yields clear predictions for how the optimal level and sensitivity of pay varies with the CEO's tenure and the contracting environment. The contract can be implemented by a "Dynamic Incentive Account": the CEO's expected pay is escrowed into an account, a proportion of which is invested in the firm's stock and the remainder in cash. The account features state-dependent rebalancing and time-dependent vesting. If the stock price falls, cash in the account is used to buy additional shares. Unlike the repricing of options, this re-incentivization does not come for free and so the CEO is not rewarded for failure. 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, private saving, myopia, short-termism, 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 an anonymous referee, an associate editor, the Editor (Cam Harvey), Raj Aggarwal, Gary Becker, Gilles Chemla, Ingolf Dittmann, Phil Dybvig, Oliver Hart, Ken Feinberg, Mike Fishman, Zhiguo He, Marcus Opp, Tomasz Piskorski, Ailsa Roell, Leonid Spesivtsev, Eric Talley, and seminar participants at the AEA, Conference on Financial Economics and Accounting, Financial Research Association, FIRS, Harvard Law School / Sloan Foundation Conference on Corporate Governance, Jackson Hole Finance Conference, LSE FMG Conference on Managers, Incentives and Organizational Structure, NBER, Paris Corporate Finance Conference, Society for Economic Dynamics, Washington University Conference on Corporate Finance, WFA, Arizona State, Caltech, Chicago, Colorado, Harvard, Northwestern, NYU, Toulouse and Wharton. Qi Liu and Andrei Savotchkine provided excellent research assistance. This paper was previously circulated under the title "Dynamic Incentive Accounts."

1 Introduction

Many classical models of CEO compensation consider only a single period, or multiple periods but a single terminal consumption. However, the optimal static contract may be ineffective in a dynamic world. In reality, 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. The CEO may be able to engage in private saving, to achieve a higher future income than intended by the contract, in turn reducing his effort incentives. Single-period contracts can encourage the CEO to engage in short-termism, i.e. inflate the current stock price at the expense of long-run value. In addition to the above challenges, a dynamic setting provides opportunities to the firm – it can reward effort with future rather than current pay.

This paper analyzes a dynamic model that allows for all of the above complications, which are likely important features in reality. We take an optimal contracting approach that allows for fully history-dependent contracts without restrictions to particular contractual forms. The key challenge of a dynamic setting with risk aversion, private saving and short-termism is that the contract is typically very complex and can only be solved numerically, which makes it difficult to see the intuition and understand which features of the setting are driving which aspects of the contract. Our main methodological contribution is to achieve a surprisingly simple optimal contract. The model's closed form solutions allow the economic forces behind the contract to be transparent, its economic implications to be clear, and a simple practical implementation using the standard instruments of cash and stock.

In the full model, the CEO engages in effort, private saving and short-termism, and the contract must achieve incentive compatibility in all three actions. The model's tractability allows us to see clearly the effect of switching these actions on and off, and thus isolate the role that each plays in determining the contract. In the simplest model, the CEO chooses only effort.

In the optimal contract, log pay is a linear function of current and all past stock returns. Therefore, the rewards for exerting effort to increase the current return are spread over all future periods, to achieve intertemporal risk-sharing. The return in any given period affects log pay in all future periods to the same degree – the first-period return has the same effect on second-period log pay as it does on first-period log pay. Moreover, in an infinite-horizon model, this sensitivity is constant across periods. Log pay is affected by returns in all past periods to the same degree – the first-period return and the second-period return have the same effect on second-period log pay. With a finite horizon, the sensitivity is increasing over time, as found empirically by Gibbons and Murphy (1992). Log pay is more sensitive to current than past returns, and the sensitivity to the current return intensifies as the CEO becomes older. This is because there are fewer periods over which to spread the reward for effort, and so the reward in the current period must increase. This effect is absent in the infinite-horizon model. We thus generate a similar prediction to the model of Gibbons and Murphy, but without invoking

career concerns.

When the CEO has the possibility of private saving, the contract must remove his incentives to undo the contract by doing so. Even if his compensation were flat, he would have a motive to save if his own level of impatience differs from that of the aggregate economy, as the latter determines the interest rate. Moreover, the presence of incentive compensation exposes him to risk which he may wish to insure against. We show that, while the contract sensitivity is affected by the model horizon, it is unaffected by whether the CEO can save privately. Instead, the possibility of private saving affects the level of pay, causing it to increase faster over time. Rising pay effectively saves for the CEO, removing the incentive for him to do so privately. That the wage should rise with tenure provides a potential explanation for seniority-based pay, which differs from existing explanations based on internal labor markets. The growth rate of consumption is increasing in the level of incentives and is thus faster for CEOs with stronger incentives (e.g. due to more severe agency problems), and accelerates over time in a finite model where incentives rise over time. However, while the possibility of private saving affects the level of pay, it has no affect on the strength of incentives.

We finally allow the CEO to engage in short-termism / myopia, e.g. by changing accounting policies or scrapping positive-NPV projects. The contract must change in several ways to prevent such behavior. When myopia is infeasible, the CEO's post-retirement income is independent of firm performance after departure, since he cannot affect it. When myopia is feasible, he can now affect post-retirement returns by engaging in short-termism prior to departure. Thus his post-retirement income must become sensitive to firm returns, to deter such actions. In addition, the contract sensitivity now rises over time, even in an infinite-horizon model. The CEO benefits from short-termism as it boosts current pay, but the cost is only suffered in the future and thus has a discounted effect. An increasing sensitivity offsets the effect of discounting by ensuring that the CEO loses more dollars in the future than he gains today. The rate of this increase and the extent of the CEO's exposure to returns after retirement are greater if the CEO is more impatient. Moreover, these direct changes to the sensitivity of the contract further induce indirect changes to the level of pay. As the sensitivity rises to deter myopia, the CEO is exposed to greater risk, in turn requiring higher pay to compensate.

The optimal contract can be implemented in the following simple manner. When appointed, the CEO is given a "Dynamic Incentive Account" ("DIA"): a portfolio of which a given fraction is invested in the firm's stock and the remainder in (interest-bearing) cash. Mathematically, the fraction of pay in stock equals the dependence of log pay on the stock return, and so it represents the contract's sensitivity. As time evolves, and firm value changes, this portfolio is constantly rebalanced to ensure the fraction of stock remains sufficient to induce effort at minimum risk to the CEO. A fall in the share price reduces the equity in the account below the required fraction; this is addressed by using cash in the account to purchase stock. If the stock appreciates, some equity can be sold without falling below the threshold, to reduce the CEO's risk.

The following numerical example illustrates the role of rebalancing. The CEO is considering whether to voluntarily forgo one week's annual leave to work on a project that will increase firm value by 10%, or take his entitled holiday which is worth 6% of his salary to him. (The higher the salary, the more the holiday is worth since he can spend his salary on holiday.) If salary is \$10m, the holiday is worth \$600,000. If the CEO has \$6m of stock, working will increase its value by 10%, or \$600,000, thus deterring the holiday. Therefore, his \$10m salary will comprise \$6m of stock and \$4m of cash. Now assume that the firm's stock price has suddenly halved, so that his stock is worth \$3m. His total salary is \$7m and the holiday is worth \$420,000, but working will increase his \$3m stock by only \$300,000. To induce effort, the CEO's gains from working must be \$420,000. This requires him to have \$4.2m of stock, and is achieved by using \$1.2m of cash in the account to purchase new stock. Importantly, the \$1.2m additional equity is not given to the CEO for free, but accompanied by a reduction in cash to \$2.8m. This addresses a concern with the current practice of restoring incentives after stock price declines by repricing options – the CEO is rewarded for failure.

The DIA also features gradual vesting: the CEO can only withdraw a percentage of the account in each period. This has three roles. First, it achieves consumption smoothing. Second, it addresses the effort problem in future periods, by ensuring that the CEO has sufficient equity in the future to induce effort. These two roles exist even if short-termism is not feasible, and requires vesting to be gradual during the CEO's employment. Third, it addresses the myopia problem in the current period, by preventing the CEO from inflating earnings and cashing out. This role requires vesting to be gradual even after the CEO retires. Gradual vesting is a more effective solution to short-termism than the clawbacks recently proposed. Clawbacks are a "cure" to recoup compensation that was paid out prematurely; gradual vesting achieves "prevention" of the premature payouts in the first place. While the former requires an explicit decision by the board and is costly to implement, the latter allows the contract to run on auto-pilot and requires no board involvement after the contract is set up.

In sum, the DIA has two key features, which each achieve separate objectives. State-dependent rebalancing ensures that the CEO always exerts effort in the current period. Time-dependent vesting ensures that the CEO has sufficient equity in future periods to induce effort, and abstains from myopia in the current period. Critical to this simple implementation is the fact that, even though consumption depends on the entire history of returns, the ratio of consumption to promised wealth (and thus the vesting fraction) and the level of incentives (and thus the fraction of stock to which the account must be rebalanced) are both history-independent. In particular, the wealth in the account is a sufficient state variable for consumption in that period; the sequence of past returns that generated that level of wealth is immaterial.

The model thus offers theoretical guidance on how compensation might be reformed to address the problems that manifested in the recent crisis, such as short-termism and weak incentives after stock price declines. A number of commentators (e.g. Bebchuk and Fried (2004), Holmstrom (2005), Bhagat and Romano (2009)) have argued that lengthening vesting

horizons may deter myopia. We provide a theoretical framework that allows to analyze and augment these verbal arguments (in particular, showing that gradual vesting is optimal even if short-termism is not feasible). While those papers focus only on lengthening vesting horizons, the DIA is critically different as it involves not only delayed vesting but also rebalancing. Delayed vesting alone only solves the myopia problem and does not ensure that the CEO's effort incentives are replenished over time – even if the CEO must hold onto his options, they have little incentive effect if they are out of the money. Moreover, in contrast to the above verbal proposals, we formally solve for the vesting fraction in a number of cases to study the optimal horizon of incentives – in particular, it is not always the case that lengthening the vesting horizon (i.e. reducing the vesting fraction) improves efficiency. In an infinite-horizon model, the vesting fraction is constant over time, and lower if private saving is possible. The agent wishes to save to insure himself against the risk imposed by equity pay; a lower vesting fraction provides automatic saving and removes these incentives. In a finite-horizon model, the fraction is increasing over time – since the CEO has fewer periods over which to enjoy his wealth, he should consume a greater percentage in later periods.

Other theories also formally model the optimal vesting horizon. The critical difference is that, in those papers, vesting and rebalancing are the same event – the CEO can only sell his securities (i.e. rebalance his portfolio) when they vest. Those papers point out that early vesting is sometimes desirable – in Chaigneau (2009) and Peng and Roell (2009), it allows the CEO to reduce his risk by trading his stock for cash; in Bhattacharyya and Cohn (2009) and Brisley (2006), this risk reduction encourages the CEO to take efficient risky projects. Thus, there is a trade-off between the benefits of early rebalancing and the costs of early vesting. In the first three papers, firms choose short-vesting stock to permit early rebalancing, even though it leads to some myopia. 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 efficient. Since rebalancing and vesting are the same event in Brisley (options can only be sold when they vest), this requires state-dependent vesting. Indeed, Bettis et al. (2010) document that performance-based (i.e. state-dependent) vesting is increasingly popular, where vesting is accelerated upon high returns.[1] However, this may induce the CEO to inflate the stock price (an action not featured in Brisley) and cash out. Here, vesting and rebalancing are separate events, allowing risk reduction without inducing myopia. High returns permit sales of equity (i.e. rebalancing) but critically the proceeds remain in the account (vesting is not accelerated) in case the returns are subsequently reversed. Our framework uses two separate instruments – vesting and rebalancing – to achieve the two separate goals of inducing effort and deterring myopia without any trade-off.

This paper is related to the dynamic agency literature, such as DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), He (2009a), Sannikov (2008), Biais et al. (2007, 2010)

[1] 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.

and Garrett and Pavan (2009, 2010). The optimal contract in these papers is typically highly complex (except if risk-neutrality is assumed, in which case private saving is a non-issue), and they do not incorporate short-termism. Lacker and Weinberg (1989), Goldman and Slezak (2006), Peng and Roell (2009) and Hermalin and Weisbach (2009) study short-termism (in the form of manipulation) in a static setting. To our knowledge, He (2009b) is the only other dynamic model featuring effort, myopia and private saving. His setup requires a discrete action choice and linear cost functions, private borrowing is ruled out, and the contract can only be solved numerically. This paper considers a fairly general setting featuring all three decisions, yet still obtain a closed-form contract which allows clear economic intuition and simple implementation. We do so by using the framework of Edmans and Gabaix (2011a) ("EG") which allows us to deliver closed-form contracts in a multi-period setting; however, EG restrict the CEO to consume in the final period only and thus cannot study private saving or short-termism, nor do they consider how to implement the contract. Holmstrom and Milgrom (1987) similarly have only terminal consumption. Allowing for intermediate consumption significantly complicates the problem. If the agent cannot save privately, the principal must solve for how to redistribute payments optimally over time to minimize the cost, creating extra optimality conditions. If the agent can save privately, the principal must solve for how to deter him from redistributing consumption to time periods with higher marginal utility, creating extra constraints.

That the optimal contract exhibits memory (i.e. current pay depends on past output) was first derived in Lambert (1983) and Rogerson (1985), who consider a two-period model where the agent only chooses effort. We extend it to a multi-period model where the agent can also save and inflate earnings. Moreover, the execution of the contract through an incentive account and thus wealth- rather than pay-based compensation allows a memory-dependent contract to be implemented simply. Bolton and Dewatripont (2005) note that a "disappointing implication of [memory-dependence] is that the long-term contract will be very complex," which appears to contradict the relative simplicity of real-life contracts. This complexity is indeed unavoidable if the CEO is rewarded exclusively through new flows of pay, as these flows will have to depend on the entire history of past outcomes.[2] Importantly, our contract can be implemented with a wealth-based account rather than with flow pay. A fall in the share price reduces the CEO's wealth and thus his entire path of future consumption. Future consumption is thus sensitive to past returns without requiring new flows of pay to be history-dependent.

In allowing for private saving, the paper makes an additional methodological contribution. To our knowledge, it is the first to derive sufficient conditions to guarantee the validity of the first-order approach to solve a multi-period agency problem with private saving and borrowing.[3] The first-order approach replaces the agent's incentive constraints against complex multi-period

[2] While long-term incentive plans (LTIPs) are used in practice and relatively simple, they typically depend on only a few years of performance rather than the entire history of performance as suggested by the model.

[3] Abraham, Koehne and Pavoni (2010) provide sufficient conditions for the first-order approach with private saving and borrowing in a two-period model, but these conditions are not sufficient for more than two periods.

deviations with weaker local constraints (first-order conditions), with the hope that the solution to the relaxed problem satisfies all constraints.[4] This method is often valid if private saving is impossible (hence the one-shot deviation principle), but problematic when the agent can engage in joint deviations to save and shirk. This is because saving insures against future shocks to income and thus reduces effort incentives. Our technique involves linearizing the agent's utility function and showing that, if the cost of effort is sufficiently convex, the linear utility function is concave in leisure (it is automatic that there is no incentive to save under linear utility). Since the actual utility function is concave, linearized utility is 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. This technique may be applicable in other agency theories to verify the sufficiency of the first-order approach.

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 and autocorrelated noise. It also provides a full justification of the contract: it derives sufficient conditions that ensure that the agent will not undertake global deviations, and shows that the principal does not want to implement a different effort level. Section 5 extends the model to allow for myopia, and Section 6 concludes. The Appendix A contains main proofs, and the Internet Appendix contains further peripheral material.

2 The Core Model

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

$$D_\tau = X \exp\left(\sum_{t=1}^{\tau} (a_t + \eta_t)\right), \tag{1}$$

where X represents baseline firm size and $a_t \in [0, \bar{a}]$ is the agent's action ("effort"). The action a_t is broadly defined to encompass any decision that improves firm value but is personally costly to the manager. Low a_t can refer to shirking, diverting cash flows or extracting private benefits. η_t is noise, which is independent across periods, has a log-concave density, and is bounded above and below by $\underline{\eta}$ and $\bar{\eta}$. (Section 4.1 allows for autocorrelated noises).

The goal of this paper is to achieve a tractable contract in a dynamic setting, to allow clear implications. Holmstrom and Milgrom (1987) show that tractability can be obtained under the joint assumptions of exponential utility, a financial cost of effort, continuous time and Gaussian

[4] Another method of verifying the validity of the first-order approach is to verify global incentive compatibility of each individual solution numerically rather than finding conditions on primitives that ensure validity. For example, see Werning (2001), Dittmann and Maug (2007) and Dittmann and Yu (2010). See also Kocherlakota (2004) for the analytical challenges of dynamic agency problems with private savings.

noise. We wish to allow for general noise distributions, decreasing absolute risk aversion (given empirical evidence), discrete time (for clarity) and non-financial effort costs. Many actions do not involve a monetary expenditure; moreover, as we will discuss, a multiplicative rather than financial cost of effort is necessary to generate empirically consistent predictions. We thus use the framework of EG who achieve tractability without the above assumptions by specifying that, in each period t, the agent privately observes η_t before choosing his action a_t. This timing assumption forces the incentive constraints to hold state-by-state (i.e. for every possible realization of η_t) and thus tightly restricts the set of admissible contracts, leading to a simple solution to the principal's problem.[5] The timing is also featured in models where the CEO sees total output before deciding how much to divert (Lacker and Weinberg (1989), DeMarzo and Fishman (2007), Biais et al. (2007)), and where the CEO observes the "state of nature" before choosing effort (Harris and Raviv (1979), Sappington (1983) and Baker (1992), and Prendergast (2002)). Note that it does not render the CEO immune to risk – in every period, except the final one, his action is followed by noise. Appendix B shows that the contract has the same form in continuous time, where η and a are simultaneous.

After action a_t is taken, the principal observes a public signal of firm value, 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.

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

With a slight abuse of terminology, we call r_t the firm's "return".[6] By observing S_t, the principal learns r_t, but not its 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 noise and the return history. After S_t (and thus r_t) is publicly observed, the principal pays the agent y_t. We allow for a history-dependent contract in which pay $y_t(r_1, \ldots r_t)$ depends on the entire history of returns.[7]

[5] Edmans and Gabaix (2011b) use this framework to achieve tractability in a market equilibrium model of CEO compensation under risk aversion.

[6] 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 innovation in the stock return is the *unexpected* increase in the stock price. In turn, the stock price is the discounted expected dividend and includes the expected future effort levels. Assuming zero risk premium for simplicity, the stock price is thus:

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

where R is the risk-free rate. Therefore, the firm's actual log return is $\ln P_t - \ln P_{t-1} = r_t - a_t^* + R - \ln E[e^{\eta_t}]$.

[7] A fully general contract can involve the income y_t depending on messages sent by the agent regarding η_t. We later derive a sufficient condition under which the optimal contract implements a fixed action, $\overline{a}$, in every

Having received income y_t, the agent consumes c_t and saves $(y_t - c_t)$ at the continuously compounded risk-free rate R. The agent may borrow as well as save, i.e. $(y_t - c_t)$ may be negative. Such borrowing and saving are unobserved by the principal. Following a standard argument (see, e.g., Cole and Kocherlakota (2001)), we can restrict attention to contracts in which the agent chooses not to save or borrow in equilibrium, i.e. $c_t = y_t$.[8] Any contract in which the CEO chooses to save to achieve a different consumption profile can be replaced by an equivalent contract providing the same consumption profile directly, so there is no loss of generality in focusing on contracts in which there is no private saving. Note this means that (as is standard) we are only uniquely solving for the agent's consumption profile, not his income profile. It could be that the principal could implement the same consumption profile with a different income profile, and the agent would voluntarily choose to save away from this income profile to achieve exactly the consumption profile intended by the agent.

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

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

where $g(a) = -\ln h(a)$, the utility cost of taking action a, is an increasing, convex function. u is a CRRA utility function with relative risk aversion coefficient $\gamma > 0$, i.e. $u(x) = x^{1-\gamma}/(1-\gamma)$ if $\gamma \neq 1$, and $u(x) = \ln x$ for $\gamma = 1$.

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.[9] 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), effort has a multiplicative effect on both CEO utility (equation (3)) and firm earnings (equation (1)). Multiplicative preferences ($u(c_t, a_t) = u(c_t h(a_t))$) 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: in particular, they are necessary for labor supply to be constant over time as wages rise; with additive preferences, leisure falls to zero as

period. Hence, on the equilibrium path, there is a one to one correspondence between r_t and η_t, which makes messages redundant: see EG for a formal proof. We allow the contract to depend on messages when providing the optimality of a fixed target action in Section 4.3. Similarly, we restrict the analysis to deterministic contracts; EG show that assuming that noise has a log-concave distribution (in addition to non-increasing absolute risk aversion, which we have) is sufficient to rule out stochastic contracts.

[8] As is standard, the CEO can save in the risk-free rate but not the stock, otherwise the CEO would be able to undo the contract and give himself a flat salary. Insider trading is illegal in nearly all countries.

[9] This assumption means that $a_t = \bar{a}$ for $t > L$. However, it could easily be weakened. The stock return after the CEO's retirement is driven only by deviations in the successor's effort level from the market's expectations (plus noise), so any publicly observed contract would have the same effect.

the wage increases.[10] With a multiplicative production function, the dollar benefits of working are higher for larger firms. Under the literal interpretation of a as effort, initiatives can be "rolled out" across the entire firm and thus have a greater effect in a larger company; under the interpretation of cash flow diversion, a large firm has more resources to steal.[11] The manager thus has a linear effect on the firm's stock return. Edmans et al. show that multiplicative specifications are necessary to deliver empirically consistent predictions for the scaling of various incentive measures with firm size.

The principal is risk-neutral and uses discount rate R. Her objective function is thus:

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

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

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

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

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

for all alternative effort strategies $(\hat{a}_t, t = 1, \ldots L)$ and *feasible* consumption strategies $(\widehat{c}_t, t = 1, \ldots T)$. 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 use the notation E^a and $E^{\hat{a}}$ to highlight that the agent's effort strategy affects the probability distribution over return paths.

The problem is complex because contracts are history-dependent, the agent can privately save, and the principal must choose the optimal effort level. Our solution strategy is as follows. We first consider a deterministic (but possibly time-varying) sequence of target actions $(a_t^*, t = 1, ..., L)$ and conjecture that the optimal contract involves binding local constraints. Following this conjecture we (i) derive the necessary local constraints that a candidate contract

[10] Bennardo, Chiappori and Song (2009) show that a multiplicative utility function can rationalize perks.

[11] See Bennedsen, Perez-Gonzalez and Wolfenzon (2009) for empirical evidence that CEOs have the same percentage effect on firm value, regardless of firm size.

must satisfy in Section 3.1; (ii) find the cheapest contract that satisfies these constraints (Theorem 1 in Section 3.2) and show that the constraints bind (Theorem 2 in Section 4.1); (iii) derive a sufficient condition under which the candidate contract is also fully incentive-compatible, i.e. prevents global deviations (Theorem 3 in Section 4.2); (iv) verify that if firm size X is sufficiently large, the optimal contract indeed involves a deterministic path of target actions: the highest effort level $a_t^* = \overline{a}$ is implemented in each period (Theorem 4 in Section 4.3).

Note that we do not require part (iv) and Theorem 4 if we wish to focus on implementing a given sequence of target actions (the first stage of Grossman and Hart (1983)) rather than also determining the optimal effort level (the second stage of Grossman and Hart). Indeed, many contracting papers focus exclusively on solving for the optimal contract to implement a given effort level, rather than jointly solving for the optimal action (see, e.g., Dittmann and Maug (2007), Dittmann, Maug and Spalt (2010)) given the substantial complexity of the latter.

3 Log Utility

3.1 Local Constraints

A candidate contract must satisfy two local constraints. The effort (EF) constraint ensures that the agent exerts the target effort level ($a_t = a_t^*$). The private savings (PS) constraint ensures that the agent consumes the full income provided by the contract ($c_t = y_t$). To highlight the effect of allowing for private savings on the contract, we also consider a version of the model in which private savings are impossible (i.e. the principal can monitor savings), and so the PS constraint is not imposed.

Consider an arbitrary contract $(y_t, t = 1, \dots T)$, a consumption strategy $(c_t, t = 1, \dots T)$ and an effort strategy $(a_t, t = 1, \dots L)$. Recall that y_t and c_t depend on the entire history $(r_1, \dots r_t)$ and a_t depends on $(r_1, \dots r_{t-1}, \eta_t)$. To capture history-dependence, E_t denotes the expectation conditional on $(r_1, \dots r_t)$. We first address the EF constraint and consider a local deviation in the action a_t after history $(r_1, \dots r_{t-1}, \eta_t)$. The effect on CEO utility is

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

Since $\partial r_t / \partial a_t = 1$ and $\partial U / \partial a_t = \rho^t c_t h'(a_t) u'(c_t h(a_t))$, the EF constraint is:

$$\begin{aligned} \text{EF} : E_t \left[\frac{\partial U}{\partial r_t} \right] &= \rho^t c_t (-h'(a_t)) u'(c_t h(a_t)), \ a_t \in (0, \overline{a}) \\ E_t \left[\frac{\partial U}{\partial r_t} \right] &\geq \rho^t c_t (-h'(a_t)) u'(c_t h(a_t)), \ a_t = \overline{a}. \end{aligned} \tag{5}$$

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.

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

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 contrasts with the "Inverse Euler Equation" (IEE), which applies to agency problems without the possibility of private saving and thus the PS constraint, when utility is additively separable in consumption and effort (e.g. Rogerson (1985) and Farhi and Werning (2009)). In our model, utility becomes additive if $u(x) = \ln x$, and the IEE is:

$$\text{IEE: } \rho^{-t}c_t = E_t\left[e^{-R}\rho^{-t-1}c_{t+1}\right]. \tag{7}$$

The inverse of the agent's discounted marginal utility $e^{-Rt}\rho^{-t}c_t$, which equals the marginal cost of delivering utility to the agent, is a martingale. If (7) did not hold, the principal would shift the agent's utility to periods with a lower marginal cost of delivering it. This argument is invalid for $\gamma \neq 1$, because the agent's marginal cost of effort depends on his consumption when utility is nonadditive.

3.2 The Contract

We now derive the cheapest contract that satisfies the local constraints. We first consider log utility as the expressions are most tractable, since the agent consumes the same amount in each period. In addition, it allows us to consider the model both with and without the PS constraint, since with log utility, the IEE applies in the case where there is no PS constraint. Section 4 considers $\gamma \neq 1$.

Theorem 1 *(Log utility.) The cheapest contract that satisfies the local constraints and implements $a_t = a_t^*\ \forall t$ is as follows. In each period t, the CEO is paid a compensation c_t which satisfies:*

$$\ln c_t = \ln c_0 + \sum_{s=1}^{t}\theta_s r_s + \sum_{s=1}^{t} k_s, \tag{8}$$

where θ_s and k_s are constants. The sensitivity θ_s is given by

$$\theta_s = \begin{cases} \frac{g'(a_s^*)}{1+\rho+...+\rho^{T-s}} \text{ for } s \leq L, \\ 0 \text{ for } s > L. \end{cases} \tag{9}$$

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

$$k_s = R + \ln \rho - \ln E[e^{\theta_s(a_s^* + \eta)}]. \tag{10}$$

If private saving is possible, k_s is given by:

$$k_s = R + \ln \rho + \ln E[e^{-\theta_s(a_s^* + \eta)}]. \tag{11}$$

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

Heuristic proof. Appendix A contains a full proof; here we present a heuristic proof in a simple case that gives the key intuition. We consider $L = T = 2$, $\rho = 1$, $R = 0$, $a_1^* = a_2^* = a^*$ and impose the PS constraint. We wish to show that the optimal contract is given by:

$$\ln c_1 = g'(a^*)\frac{r_1}{2} + \kappa_1, \quad \ln c_2 = g'(a^*)\left(\frac{r_1}{2} + r_2\right) + \kappa_1 + k_2 \tag{12}$$

for some constants κ_1 (the equivalent of $\ln c_0 + k_1$ in the Theorem) and k_2 that make 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.:

$$\ln c_1 = \theta_1 r_1 + \kappa_1, \ \ln c_2 = \theta_{21} r_1 + \theta_2 r_2 + \kappa_1 + k_2 \tag{13}$$

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

First, intuitively, the optimal contract entails consumption smoothing, i.e. shocks to consumption are permanent. This implies $\theta_{21} = \theta_1$. To prove this, the PS constraint (6) yields:

$$1 = E_1\left[\frac{c_1}{c_2}\right] = e^{(\theta_1 - \theta_{21})r_1} E_1\left[e^{-\theta_2 r_2 - k_2}\right]. \tag{14}$$

This must hold for all r_1. Therefore, $\theta_{21} = \theta_1$ and $k_2 = \ln E_1\left[e^{-\theta_2 r_2}\right]$, as in (11).

Next, consider total utility U:

$$\begin{aligned} U &= \ln c_1 + \ln c_2 - g(a_1) - g(a_2) \\ &= 2\theta_1 r_1 + \theta_2 r_2 - g(a_1) - g(a_2) + 2\kappa_1 + k_2. \end{aligned}$$

From (5), the two EF conditions are $E_2\left[\frac{\partial U}{\partial r_1}\right] \geq g'(a^*)$ and $E_2\left[\frac{\partial U}{\partial r_2}\right] \geq g'(a^*)$, i.e.:

$$2\theta_1 \geq g'(a^*), \quad \theta_2 \geq g'(a^*).$$

Intuitively, the EF constraints should bind (proven in the Appendix), else the CEO is exposed to unnecessary risk. Combining the binding version of these constraints with (13) yields (12).

Step 2: Optimality of log-linear contracts

We next verify that optimal contracts should be log-linear. Equation (5) yields: $d(\ln c_2)/dr_2 \geq g'(a^*)$. The cheapest contract involves this local EF condition binding, i.e.

$$d(\ln c_2)/dr_2 = g'(a^*) \equiv \theta_2. \tag{15}$$

Integrating yields the contract:

$$\ln c_2 = \theta_2 r_2 + B(r_1), \tag{16}$$

where $B(r_1)$ is a function of r_1 which we will determine shortly. It is the integration "constant" of equation (15) viewed from time 2.

We next apply the PS constraint (6) for $t = 1$:

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

Hence, we obtain

$$\ln c_1 = B(r_1) + K, \tag{18}$$

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

$$U = \ln c_1 + \ln c_2 + K' = \theta_2 r_2 + 2B(r_1) + 2K + K'. \tag{19}$$

We next apply (5) to (19) to yield: $2B'(r_1) \geq g'(a^*)$. Again, the cheapest contract involves this condition binding, i.e. $2B'(r_1) = g'(a^*)$. Integrating yields:

$$B(r_1) = g'(a^*)\frac{r_1}{2} + K''. \tag{20}$$

Combining (20) with (18) yields: $\ln c_1 = g'(a^*)\frac{r_1}{2} + \kappa_1$, for another constant κ_1. Combining (20) with (16) yields:

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

for some constant k_2. ■

The contract's closed-form solutions allow transparent economic implications. (8) shows that time-t income should be linked to the return not only in period t, but also in all previous

periods. Therefore, changes to r_t (due to effort or shocks) boost log pay in the current and all future periods equally. Since the CEO is risk-averse, it is efficient to spread the effect of effort and noise over the future. Indeed, Boschen and Smith (1995) find empirically that firm performance has a much greater effect on the NPV of future pay than current pay.

We now consider how the contract sensitivity changes over time. We consider the case of a fixed target action ($a_t^* = a^* \ \forall \ t$) so that the changes in the contract's sensitivity are not driven by changes in the implemented effort level. (9) shows that, in an infinite horizon model ($T = \tau \to \infty$), the sensitivity is constant and given by:

$$\theta_t = \theta = (1 - \rho)\, g'(a^*). \tag{21}$$

This is intuitive: the contract must be sufficiently sharp to compensate for the disutility of effort, which is constant. Thus, not only does r_t have the same effect on log consumption in every period, but also $\ln c_t$ is affected by the return in every period to the same degree. The sensitivity to the current-period return is decreasing in the discount rate – if the CEO is more impatient (lower ρ), it is necessary to reward him today rather than in the future.

However, for any model with finite life T, (9) shows that θ_t is increasing over time. To understand the intuition for this increasing sensitivity, we distinguish between the increase in lifetime utility for exerting effort ($\partial U/\partial a_t$) and the increase in current utility ($\partial u_t/\partial a_t = \theta_t$); the latter also equals the increase in current log consumption ($\partial \ln c_t/\partial a_t$). Since the disutility of effort is constant, the lifetime utility reward for effort, $\partial U/\partial a_t$, must also be constant. When there are fewer remaining periods over which to smooth out this lifetime increase, the increase in current utility ($\partial u_t/\partial a_t$) must be higher. By contrast, Gibbons and Murphy (1992) generate an increasing current sensitivity because the lifetime increase in utility $\partial U/\partial a_t$ rises over time, to offset falling career concerns. In Garrett and Pavan (2009), the current sensitivity rises over time because $\partial U/\partial a_t$ increases to minimize the agent's informational rents. Here, $\partial U/\partial a_t$ is constant since we have no adverse selection or career concerns; instead the increase in $\partial u_t/\partial a_t$ stems from the reduction in consumption smoothing possibilities as the CEO approaches retirement. Both Gibbons and Murphy (1992) and Cremers and Palia (2010) document that incentives increase with CEO tenure.

As in the infinite-horizon case, the sensitivity to the current return decreases with discount rate ρ. In the finite-horizon case, ρ also determines the speed at which incentives rise over time. If the CEO is more patient, the contract involves greater consumption smoothing to begin with, and so is more greatly affected by the decline in consumption smoothing possibilities as retirement approaches. Thus, incentives increase particularly rapidly for more patient CEOs.

While θ_t depends on the model horizon, it is independent of whether private saving is possible – this only affects k_t. Since private saving does not affect the agent's action and thus firm returns, the sensitivity of pay to returns is unchanged. Instead, it alters the time trend in the level of pay. The log expected growth rate in pay is, from (8): $\ln E\left[c_t/c_{t-1}\right] = k_t + \ln E\left[e^{\theta_t r_t}\right]$.

If private saving is impossible, substituting for k_t using (10) yields:

$$\ln E\left[c_t/c_{t-1}\right] = R + \ln \rho,$$

which is constant over time and independent of risk. The risk-free rate R is determined by the time preference of the aggregate economy. If and only if the CEO is more patient than the representative agent, then the growth rate is positive, as is intuitive. If private saving is possible, (11) yields:

$$\ln E\left[c_t/c_{t-1}\right] = R + \ln \rho + \ln E[e^{-\theta_t r_t}] + \ln E[e^{\theta_t r_t}].$$

In the limit of small time intervals (or, equivalently, in the limit of small variance of noises σ^2), this yields:

$$\ln E\left[c_t/c_{t-1}\right] = R + \ln \rho + \theta_t^2 \sigma_t^2.$$

Thus, the growth rate of consumption is always greater where private saving is possible. This 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 (2009b), who finds that the optimal contract under private savings involves a wage pattern that is non-decreasing over time.[12] The model thus predicts a positive relationship between the wage and tenure, which is consistent with the common practice of seniority-based pay. Cremers and Palia (2010) confirm this relationship empirically. Moreover, the growth rate depends on the risk to which the CEO is exposed, which is in turn driven by his sensitivity to the firm's returns θ, and the volatility of firm returns σ. CEOs with stronger incentives (e.g. because the agency problem is more severe) or who work in riskier firms will have pay growing more rapidly over time. This is intuitive: a rising level of pay insures the CEO from risk, removing the need for him to do so himself. Furthermore, in a finite-horizon model, θ_t is increasing over time and so the growth rate of consumption rises with tenure, i.e. pay accelerates over time.

We can also calculate how much the expected cost of compensation rises if private saving is possible and the principal must impose the PS constraint – i.e. the cost to the principal of her inability to monitor the CEO's private savings. We follow the analysis of Farhi and Werning (2009) for this calculation.

Corollary 1 *(Cost of Private Savings). Define* $\Lambda =$ *(Expected cost of contract imposing PS) / (Expected cost of contract without imposing PS), and consider* $L = T = \infty$ *and* $a_t^* = a^*\ \forall t$. *We have* $\Lambda \geq 1$ *and:*

$$\Lambda = \frac{1-\rho}{1-\rho e^{\Theta^2\sigma^2}} e^{-\frac{\rho\Theta^2\sigma^2}{1-\rho}},$$

[12] Lazear (1979) has a back-loaded wage pattern for incentive, rather than private saving considerations (the agent is risk-neutral in his model). Since the agent wishes to ensure he receives the high future payments, he induces effort to avoid being fired. Similarly, in Yang (2009), a back-loaded wage pattern induces agents to work to avoid the firm being shut down.

using the notation $\Theta^2\sigma^2 = \ln E\left[e^{-\theta\eta}\right] + \ln E\left[e^{+\theta\eta}\right]$. *In the limit of small time intervals,* $\Theta \sim \theta = (1-\rho)\, g'(a^*)$ *and* $\Lambda \sim e^{-\frac{\theta^2\sigma^2}{1-\rho}} / (1 - \theta^2\sigma^2/(1-\rho))$.

The ratio Λ increases in the risk borne by the agent, $\theta^2\sigma^2$ as this affects his desire to save. In addition, from (21) we see that Λ is closer to one when the agent is more patient.

The contract in Theorem 1 also has implications for the appropriate measure of incentives. Taking first differences of this contract yields:

$$\ln c_t - \ln c_{t-1} = \theta_t r_t + k_t. \tag{22}$$

The percentage change in CEO pay is linear in the firm's return r_t, i.e. the percentage change in firm value. Thus, the relevant measure of incentives is the percentage change in pay for a percentage change in firm value ("percent-percent" incentives), or equivalently the elasticity of CEO pay to firm value; in real variables, this equals the fraction of total pay that is comprised of stock. This elasticity/fraction must be θ_t to achieve incentive compatibility and is independent of firm size. "Percent-percent" incentives are relevant because effort has a multiplicative (i.e. percentage) effect on both CEO utility and firm value.

Empiricists have used alternative statistics to measure incentives – 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 elasticities ("percent-percent" incentives) on empirical grounds: they are invariant to firm size and thus comparable across firms of different size (as found by Gibbons and Murphy (1992)), and firm returns have greater explanatory power for percentage than dollar changes in pay. Thus, firms behave as if they target percent-percent incentives. However, he notes that "elasticities have no corresponding agency-theoretic interpretation." Our framework provides a theoretical justification for using elasticities to measure incentives. Edmans et al. (2009) show that multiplicative preferences and production functions generate elasticities as the incentive measure, which motivates their use here (equations (1) and (3)).[13] Their result was derived in a one-period model with a risk-neutral CEO; we extend it to a dynamic model with risk aversion and private saving.

The contract in Theorem 1 involves binding local constraints and implements $a_t = \bar{a}$. The remaining steps are to show that the agent will not undertake global deviations (e.g. make large single-action changes, or simultaneously shirk and save) and that the principal cannot improve by implementing a different effort level or allowing slack constraints. Since these proofs are equally clear for general γ as for log utility, we delay them until Section 4.

[13] Peng and Roell (2009) also use a multiplicative specification and restrict analysis to contracts where log pay is linear in firm returns. This paper endogenizes the contract form and thus provides a microfoundation for considering only loglinear contracts.

3.2.1 A Numerical Example

This section uses a simple numerical example to show most clearly the economic forces behind the contract. We first set $T = 3$, $L = 3$, $\rho = 1$, $a_t^* = a^*$ and $g'(a^*) = 1$. From (9), the contract is:

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

where $\kappa_t = \sum_{s=1}^{t} k_s$. An increase in r_1 leads to a permanent increase in log consumption – it rises by $\frac{r_1}{3}$ in all future periods. In addition, the sensitivity $\partial u_t / \partial a_t$ increases over time, from 1/3 to 1/2 to 1/1. The total lifetime reward for effort $\partial U_t / \partial a_t$ is a constant 1 in all periods.

We now consider $T = 5$, so that the CEO lives after retirement. The contract is now:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3 \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_4 \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_5.
\end{aligned} \tag{23}$$

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

3.3 Implementation: the Dynamic Incentive Account

The contract derived in Section 3.2 can be implemented in at least two ways. First, it can be implemented using purely flow-based pay: the principal simply pays the agent the amount c_t given by Theorem 1. Second, it can be implemented using a wealth-based account, as described in Proposition 1 below.

Proposition 1 *(Contract Implementation via a Dynamic Incentive Account). In a finite-horizon model, the contract in Theorem 1 can be implemented as follows. The present value of the CEO's expected pay is escrowed into a "Dynamic Incentive Account" ("DIA") at the start of $t = 1$.*[14] *A proportion θ_1 is invested in the firm's stock and the remainder in interest-bearing*

[14] If the CEO has any initial wealth, it is also placed in the DIA. In reality, managers of start-ups often co-invest in their firm. Note that the stock pays the firm's actual return. As noted in footnote 6, r_t is not the

cash. At the start of each subsequent period t, the DIA is rebalanced so that the proportion invested in the firm's stock is θ_t. A deterministic fraction α_t vests at the end of each period and can be withdrawn for consumption. The vesting fraction is given by:

$$\alpha_t = c_t/A_t = 1/E_t \left[\sum_{s=t}^{T} e^{-R(s-t)} e^{\sum_{n=t+1}^{s} \theta_n \eta_n + k_n} \right]$$

(i) If private saving is impossible and $a_t^* = a^* \ \forall\, t$, α_t has a particularly simple form and is given by $\alpha_t = (1-\rho) \,/\, \left(1-\rho^{T-t}\right)$.

(ii) In an infinite-horizon model in which private saving is possible, $a_t^* = a^* \ \forall\, t$, and noise η_t is i.i.d., the contract can be implemented by a DIA with $\alpha_t = \alpha = 1 - \rho E\left[e^{\theta\eta}\right] E\left[e^{-\theta\eta}\right] < 1-\rho$, as long as $\alpha > 0$.

The rebalancing of the DIA ensures that θ_t of the agent's wealth is invested in stock at all time, so that his percent-percent incentives equal θ_t. This rebalancing addresses a common problem of options: if firm value declines, their delta and thus incentive effect is reduced. Unrebalanced shares suffer a similar problem, even though their delta is 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 CEO's equity as a fraction of his wealth (percent-percent incentives). When the stock price falls, this fraction, and thus the CEO's incentives, are reduced – intuitively, when the firm becomes smaller, effort has a smaller dollar impact (given a multiplicative production function) and so a greater dollar value of stock is necessary to preserve effort incentives.

The DIA addresses this problem by exchanging stock for cash, to maintain the fraction 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.[15] On the other hand, if the share price rises, the stock fraction grows. Therefore, some shares can be sold for cash, reducing the CEO's risk, without incentives falling below θ_t. Indeed, Fahlenbrach and Stulz (2009) find that decreases in CEO ownership typically follow good performance.[16] Core and Larcker (2002) study stock ownership guidelines, whereby boards set minimum requirements for executive shareholdings. In 93% of cases, the requirements relate to the value of shares as a multiple of salary: consistent with our model, this involves rebalancing (giving additional stock after the price has fallen to

firm's actual return, but the actual return plus a constant. This does not affect the implementability with stock because it only changes the constant k_t, which rises by $\theta_t(a^* - R + \ln E\left[e^{\eta_t}\right])$.

[15] Achraya, John and Sundaram (2000) show that the cost of rewarding failure may be outweighed by the benefit of reincentivization, and so repricing options can be optimal. The rebalancing in the DIA achieves the benefit of reincentivization without the cost of rewarding failure.

[16] Fahlenbrach and Stulz (2009) measure CEO ownership by the percentage of shares outstanding (dollar-dollar incentives), rather than percent-percent incentives θ_t. Thus, ownership must fall (rise) with good (bad) performance to keep θ_t constant.

maintain a constant multiple) and implies targeting of percent-percent incentives. The idea of rebalancing incentive portfolios is similar to the widespread practice of rebalancing investment portfolios: both are ways of maintaining desired weights in response to stock price changes.

The DIA thus features dynamic rebalancing to ensure that the EF constraint is satisfied in the current 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 gradual vesting. This vesting is time-dependent: regardless of the account's value, the CEO can only withdraw a percentage α_t in each period for consumption. The fraction α_t is history-independent. This gradual vesting has two roles. First, it achieves consumption smoothing. Second, it ensures that the EF constraint is satisfied in future periods, by guaranteeing that the CEO has sufficient wealth in the account for the principal to "play with" so that she can achieve the required equity stake by rebalancing this wealth. If the CEO is allowed to fully withdraw his wealth from the account, his wealth would be outside the principal's control and so she would not be able to rebalance it. This motivation exists during the CEO's employment only – the account fully vests in period L. The CEO is not exposed to returns after period L as he cannot affect them and so any exposure would merely subject him to unnecessary risk. Note that this motivation for gradual vesting contrasts existing verbal arguments based on deterring myopic actions (e.g. Bebchuk and Fried (2004), Holmstrom (2005), Bhagat and Romano (2009)). While we show in Section 5 that allowing for such actions provides an additional case for gradual vesting, the core model demonstrates that gradual vesting is optimal even if short-termism is not possible.

Moreover, in contrast to the above verbal arguments on the vesting horizon, Proposition 1 explicitly solves for the optimal vesting rate in a number of benchmark cases. This allows us to analyze the economic forces that affect the vesting rate. If private saving is feasible and the model horizon is infinite, part (i) specializes to $\alpha = 1 - \rho$. Thus, the vesting fraction is time-independent, just like the contract sensitivity θ_t. If the horizon is finite, $\alpha_t = (1-\rho) / \left(1 - \rho^{T-t}\right)$ and is increasing over time. This is intuitive: since the CEO has fewer periods over which to enjoy his wealth, he should consume a greater percentage in later periods. Part (ii) shows that, in an infinite horizon model where private saving is possible, we have $\alpha < 1 - \rho$. The agent would like not to hold stock as it carries a zero risk premium, but is forced to invest $\theta\%$. He thus wishes to save to insure himself against this risk. To remove these incentives, we have $\alpha < 1 - \rho$ so that the account grows faster than it vests, thus providing automatic saving. In both (i) and (ii), the vesting fraction increases when the CEO is more impatient (i.e. ρ is lower), as is intuitive.

One aspect of a wealth-based implementation that we do not model explicitly is the funding of the DIA by the firm. In the simplest case of Proposition 1, the present value of the CEO's future salary is placed in the account when he is initially appointed. Alternatively, the firm may smooth out these contributions over time by funding the account gradually. In addition, the DIA (regardless of how it is funded) represents only one implementation of the contract. Other implementations are possible: rather than setting up an account and rebalancing, the

principal can simply pay the agent c_t in each period, i.e. implement the contract with purely flow compensation. The DIA implementation highlights the economic interpretation of such a payment scheme: it has the same effect *as if* the NPV of the CEO's future pay was escrowed, rebalanced and gradually vested. The interest in showing that the contract can be implemented via a wealth-based account is that this allows consumption to be history-dependent, without new flows of pay having to depend on past returns in a complex manner, as discussed in the Introduction.

4 Generalization and Justification

This section is divided as follows. Section 4.1 generalizes our contract to all CRRA utility functions and autocorrelated noise, and shows that the local EF constraint must bind. Section 4.2 derives sufficient conditions for the contract to be fully incentive compatible (i.e. deters global deviations) and Section 4.3 proves that, if the firm is sufficiently large, the optimal contract indeed involves a deterministic effort level – it requires $a_t = \bar{a}$ after every history. Section 4.4 discusses the role played by each of the assumptions in generating the model's key results.

4.1 General CRRA Utility and Autocorrelated Signals

The core model assumes that the signal r_t was the firm's stock return and so it is reasonable to assume the noises η_t are uncorrelated. However, in private firms, there is no stock return; for some public firms, the stock is illiquid and thus an inaccurate measure of performance. Therefore, 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 assume that η_t follows 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 and bounded above and below by $\underline{\varepsilon_t}$ and $\overline{\varepsilon_t}$.

We also now allow for a general CRRA utility function. Note that for $\gamma \neq 1$, the IEE is not valid if private savings are impossible, so we only consider the case where the PS constraint is imposed. We define $J_t = \rho^t e^{-(1-\gamma)g(a_t^*)}$ for $t \leq L$ and $J_t = \rho^t$ otherwise. The optimal contract is given in Theorem 2 below. Even though the principal must rule out private savings, she still has freedom in the choice of the contract (and so the optimization problem remains complex) if she wishes to implement a boundary action (Theorem 4 gives sufficient conditions under which a boundary action is optimal.) With a boundary action, the principal could use a contract with a greater sensitivity than necessary. Theorem 2 proves that this is suboptimal.

Theorem 2 *(General CRRA utility, autocorrelated noise, with the PS constraint.) The cheapest contract that satisfies the local constraints and implements $a_t = a_t^* \; \forall\, t$ is as follows. In each*

period t, the CEO is paid c_t which satisfies:

$$\ln c_t = \ln c_0 + \sum_{s=1}^{t} \theta_s \left(r_s - \phi r_{s-1}\right) + \sum_{s=1}^{t} k_s, \tag{24}$$

where θ_s and k_s are constants and $r_0 = 0$. The sensitivity θ_s is given by:

$$\theta_s = \begin{cases} \frac{J_s(g'(a_s^*) - \phi\theta_{s+1})}{\sum_{m=s}^{T} J_m \prod_{n=s+1}^{m} E_s\left[e^{(1-\gamma)\left[\theta_n(\varepsilon_n + a_n^* - \phi a_{n-1}^*) + k_n\right]}\right]} + \phi\theta_{s+1} \text{ for } s \leq L, \\ 0 \text{ for } s > L. \end{cases} \tag{25}$$

The constant k_s is given by:

$$\gamma k_s = R + \ln\rho - (1-\gamma) g(a_s^*) \mathbf{1}_{s=L+1} + \ln E\left[e^{-\gamma\theta_s(\varepsilon_s + a_s^* - \phi a_{s-1}^*)}\right] \text{ for } s \leq T. \tag{26}$$

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

If $L = T = \infty$ and $a_t^* = a^* \ \forall\ t$, the sensitivity (25) simplifies to a constant $\theta_s = \theta$, where θ is given by (44) in Appendix A. In the limit of small time intervals, and when $\phi = 0$, we have:

$$\theta = \frac{1 - \sqrt{1 - 2(\gamma-1)\sigma^2 g'(a^*)^2 \frac{(\gamma-1)R - \ln\rho}{\gamma}}}{(\gamma-1)\sigma^2 g'(a^*)}. \tag{27}$$

and $k_s = k = (R + \ln\rho)/\gamma - \theta a^* - \gamma\theta^2\sigma^2/2$.

Equation (24) shows that moving from log to general CRRA utility but retaining independent noise has little effect on the functional form of the optimal contract, which remains in closed-form and independent of the noise distribution. Similarly, γ only affects the specific values of θ and k rather than the functional form. The time trend of the contract sensitivity and the implementation via the DIA remain the same. The difference is that the parameters θ and k are somewhat more complex. To understand the economic forces that determine θ, consider the benchmark case where $\phi = 0$, $L = T$, and $a_t^* = a^*$, $\forall t$. Then, the sensitivity (25) becomes

$$\theta_t = \frac{J_t c_t^{1-\gamma}}{\sum_{s=t}^{T} E\left[J_s c_s^{1-\gamma}\right]} g'\left(a^*\right), \tag{28}$$

which stems directly from the EF condition. Under plausible parameterizations of the model (e.g., small time intervals, or $\ln\rho + R$ is close to 0), when $\gamma \geq 1$, the sensitivity increases over time up to $\theta_T = g'\left(a^*\right)$ and is steeper if the agent is more risk averse (higher γ) and less patient (lower ρ), and stock return volatility is higher. (The full derivations are in Appendix C.) Intuitively, these changes decrease the utility the agent derives from future consumptions, $\sum_{s=t}^{T} E\left[J_s c_s^{1-\gamma}\right]$, which is in the denominator of (28). Since future rewards are insufficient to induce effort, the CEO must be given a higher sensitivity to current consumption.

Equation (24) shows that, with autocorrelated signals, the optimal contract 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. Similarly, if there is an industry-wide component to r_t, the optimal contract will filter out this component, just as it filters out ϕr_{t-1}. Thus, relative performance evaluation can be combined with the contract.

4.2 Global Constraints

We have thus far derived the best contract that satisfies the local constraints. We now verify that this contract also satisfies the global constraints, i.e. the agent will not undertake global deviations. The following analysis derives a sufficient condition on g to guarantee this.

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

$$\ln y_t = \ln c_0 + \sum_{s=1}^{t} \theta_s(a_s + \eta_s - \phi(a_{s-1} + \eta_{s-1})) + \sum_{s=1}^{t} k_s, \tag{29}$$

The following Theorem states that if the cost function g is sufficiently convex and the target effort level does not rise too rapidly over time, the CEO has no profitable global deviation.

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

$$\max_{a_t, c_t \ adapted} E\left[\sum_{t=1}^{T} \rho^t u\left(c_t e^{-g(a_t)}\right)\right] \tag{30}$$

with $\sum_{t=1}^{T} e^{-rt}(y_t - c_t) \geq 0$ *and* y_t *satisfying (29). If function* g *is sufficiently convex (i.e.* $\inf_a g''(a)$ *is sufficiently large) and* $\theta_t - \phi\theta_{t+1} \geq 0 \ \forall \ t$*, the solution of this problem is* $c_t \equiv y_t$*,* $t \leq T$*, and* $a_t = a_t^*$*,* $\forall t$*. There is no global deviation from the recommended policy that makes the agent better off.*

The role of the condition on the convexity of the cost function is standard. The intuition for the condition that $\theta_t - \phi\theta_{t+1} \geq 0 \ \forall \ t$ is that, if the target effort level (and thus contract slope θ_t) rises rapidly over time, the agent will shirk in period t. This will reduce the period t return r_t and thus his consumption c_t, but increase his wage in all future periods – if noise is highly autocorrelated (ϕ is high), then the combination of a low r_t and high returns in future periods will fool the principal into thinking that the agent exerted higher effort in periods $t+1$ onwards than he actually did. Formally, the problem becomes non-concave. Thus, we require either low autocorrelation in the noise (low ϕ) or the target action not to rise too rapidly over time. Indeed, in Theorem 4 we show that, if firm size X is sufficiently large, the optimal contract involves a constant effort level.

The proof, in the Appendix, may be of general methodological interest. It involves three steps. First, we reparameterize the agent's utility from a function of consumption and effort to one of consumption and leisure, where the new variable, leisure, is defined so that utility is jointly concave in both arguments. Second, we construct an "upper-linearization" function: we create a surrogate agent with a linear state-dependent utility. Third, we prove that any global deviation by the surrogate agent weakly reduces his utility. It is automatic that there is no motive to save under linear utility. Turning to effort, if the cost of effort g is sufficiently convex,[17] the PV of the agent's income is concave in leisure. Since utility is linear in consumption, and consumption equals income, utility is concave in leisure and so there is no profitable deviation. Since our original agent's utility function is concave, his utility is the same as the surrogate agent's under the recommended policy, and weakly lower under any other policy. Thus, any deviation also reduces the original agent's utility. The third step is a Lemma that shows that the PV of income is a concave function of actions under suitable reparameterization. It thus may have broader applicability to other agency theories, allowing the use of the first-order approach to significantly simplify the problem.

4.3 The Optimality of High Effort

This section derives conditions under which the principal wishes to implement the boundary effort level $a_t = \bar{a}$ in every period and after every history. We refer to $\overline{a}$ as "high effort", to use similar terminology to models with discrete effort levels (e.g. high, medium, low) in which the high effort level is typically optimal.

Theorem 4 *(High effort is optimal if the firm is sufficiently large.) Assume that* $\inf_{\eta\in(\underline{\eta},\overline{\eta})} f(\eta) > 0$ *and* $\sup_{a\in(\underline{a},\overline{a})} g''(a)/g'^2(a) < \infty$*, where* f *is the probability density of* η*. There exists* X_* *such that if baseline firm size* $X > X_*$*, implementing* $a_t = \bar{a}$ *is optimal.*

The intuition is as follows. For any alternative contract satisfying the incentive constraints, we compare the benefits and costs of moving to a high effort contract. The benefits are multiplicative in firm size. The costs comprise the direct disutility from working, the risk premium required to compensate the CEO for a variable contract, and the change in CEO's informational rent (which are all a function of the CEO's wage). Since the CEO's wage is substantially smaller than firm size, the benefits of high effort outweigh the costs. In practice, a boundary effort level arises because there is a limit to the number of productive activities the CEO can undertake to benefit the principal. Under the literal interpretation of a as effort, there is a finite number of positive-NPV projects available and a limit to the number of hours a day the

[17] See Dittmann and Yu (2010) for a similar convexity condition to ensure that the local optimum is globally optimal. They consider a one-period model where private savings are not possible, but the CEO chooses risk as well as effort.

CEO can work while remaining productive. Under the interpretation of a as rent extraction, $\overline{a}$ reflects zero stealing.

The complexity in the proof lies in deriving an upper bound on the informational rent (which stems from the CEO's private information about the noise η) and the risk imposed on the CEO from incentives (which depends on the CEO's ability to self-insure via privately saving). Any change in the implemented effort level requires adjusting the wage not only in a particular period for the whole range of noises, but also across time periods to deter private saving. Implementing $a_t = \bar{a}$ in period t requires the time-t contract to change. Moreover, the change in the time-t contract has a knock-on effect on the time $t-1$ contract, which must change to deter saving between time $t-1$ and time t. The change in the time $t-1$ contract impacts the time $t-2$ contract, and so on: due to private saving, the contract adjustments "resonate" across all time periods. It is this non-separability which significantly complicates the problem. These complications are absent in EG, who derive a similar result in a single-period model.

This above result may be of use for future theories by simplifying the contracting problem. Grossman and Hart (1983) solved the one-period contracting problem in two stages: finding the cheapest contract that implements a given effort level, and then finding the optimal effort level. Solving both stages is typically highly complex; indeed, Grossman and Hart can only do so numerically. The idea that the benefits of effort are orders of magnitude higher than the costs simplifies the problem – since high effort is optimal, the second stage of the contracting problem is solved and so the analysis can focus exclusively on the first stage.

4.4 Discussion of Modeling Assumptions

This subsection discusses which of the model's assumptions are necessary for its key results. We view the paper's main contributions as threefold:

E. (Economic): Economic insights on the forces that drive the optimal contract, e.g. how the sensitivity θ_t and level k_t of pay change over time and depend on the environment; how the CEO remains exposed to firm returns after retirement if short-termism is possible.

T. (Tractability): Achieving a simple, closed-form optimal contract in a dynamic setting with private saving and short-termism.

I. (Implementation): The contract can be implemented with a wealth-based account, with state-dependent rebalancing and time-dependent vesting (I1). The account contains the standard instruments of stock and cash (I2)

Note that (E) and (I) are distinct implications. The contract in Theorem 1 can always be implemented with flow pay, i.e. paying the CEO an amount c_t in every period, and all the economic implications of the contract would follow. (I) refers to only one simple implementation.

We now discuss the roles played by the main assumptions in generating the above results:

A1. *CRRA utility and multiplicative preferences.* We consider these assumptions together as they are closely intertwined – the former (latter) means that an agent's allocation to risky assets (leisure) is proportional to his wage. EG show that these assumptions are not necessary for a simple contract if there is only terminal consumption. However, they are important in a model with intermediate consumption as they lead to multiplicative separability and key variables scaling with the wage. To understand the importance of multiplicative preferences for (T), assume $L = T$ and consider the final period L. With multiplicative preferences, the incentive measure is the elasticity of pay to firm value. This elasticity must be θ_L, irrespective of the level of pay in period L – and is thus *independent of the history of past returns*. The principal can thus defer the rewards for performance in prior periods (to smooth consumption) without distorting effort incentives. Deferral affects the level of pay in period L but not effort incentives, as long as the elasticity remains θ_L.

Multiplicative preferences also mean that the whole promised wealth of the agent in period 1 is multiplicative in $c_1(r_1)$, promised wealth at period 2 conditional on r_1 is multiplicative in $c_2(r_1, r_2)$[18], and so on. In other words, a shock to r_1 has a multiplicative effect on consumption in all future periods. Moreover, when we also have CRRA utility, this multiplicative effect is *the same in every future period*, for optimal risk-sharing. If r_1 falls by 2%, log consumption falls by $C \times 2\%$ in the current and all future periods, where C is a constant. Thus, rewards for performance are smoothed in a simple manner, and this smoothing is also *independent of the history of past returns* – for example, the effect of r_2 on $c_2, ..., c_T$ is independent of r_1.[19] Together, both assumptions mean that, although consumption is history-dependent, θ_t is history-independent and so the dynamic contract is a simple extension of the static contract.

The assumptions also allow a wealth-based implementation, i.e. (I1). Since wealth is a multiple of consumption, consumption is a fraction of wealth. We can therefore implement the contract by investing the CEO's wealth into instruments that yield $c_1(r_1)$ in the first period, allowing him to consume a fraction α_1, then rebalancing by investing the remainder of his wealth in instruments that yield $c_2(r_1, r_2)$ as a function of r_2, and so on. The thresholds to which the account must be rebalanced θ_t are history-independent, since the elasticity is history-independent. Furthermore, since the return in a particular period has the same effect on all future consumptions, the ratio of current consumption

[18] We require $c_2(r_1, r_2) = c_1(r_1) f(r_2)$, $c_3(r_1, r_2, r_3) = c_2(r_1, r_2) g(r_2)$ etc., i.e. multiplicative separability.

[19] With multiplicative preferences but without CRRA, the smothing is complex and history-dependent. Consider a 2-period model with $u(c, a) = e^{ch(a)}$. We have $c_2(r_1, r_2) = B(r_1)e^{\theta_2 r_2}$, and PS yields $e^{c_1 h(\overline{a})} = E_1\left[e^{B(r_1)e^{\theta_2 r_2} h(\overline{a})}\right]$. Even though r_1 has a multiplicative effect on c_2, solving for the magnitude of this effect $B(r_1)$ is highly complex.

to the sum of all future consumptions (i.e. wealth) is a constant and is independent of past shocks. Thus, the CEO's promised wealth is a sufficient statistic for his current consumption – the sequence of past returns that led to the CEO accumulating this level of wealth is irrelevant. Since consumption depends on current wealth alone, the vesting fraction α_t is history-independent.[20]

Multiplicative separability is not necessary for (T) – additive separability with CARA utility and additive preferences would also work (see Appendix E); the above arguments apply but with dollar amounts replacing percentage amounts. However, the model would predict that dollar-percent incentives are the relevant measure and independent of pay and firm size (contrary to evidence, e.g. Jensen and Murphy (1990)). Moreover, it would not permit a wealth-based implementation, i.e. (I1). With multiplicative preferences, the relevant measure of incentives is percent-percent incentives, which equals the fraction of wealth that is in stock. Regardless of the level of wealth, it can always be rebalanced to ensure that the fraction is at the required level. By contrast, dollar-percent incentives equal the dollar value of equity. If the value of the account falls below the required dollar equity holding, there is no way that it can be rebalanced to restore the CEO's equity holdings to this threshold, since cash cannot be negative owing to limited liability. Put differently, if a fall in returns reduces future consumption by a fixed dollar amount, after sufficiently many periods of low returns, the required future consumption would be negative.

Multiplicative preferences are not necessary for (E). In any model with myopia, the CEO must remain tied to firm returns after he retires. The time trend in θ is determined by consumption smoothing motives and the time trend in k is determined by the need to save for the agent; neither hinge on the specific preference formulation.

A2. *Multiplicative production function.* This assumption is used in the proof of the optimality of $a_t = \overline{a}$ in Theorem 4. It is a sufficient, rather than necessary condition for this result – as long as the dollar benefits of effort are increasing in (although not necessarily proportional to) firm size, $a_t = \overline{a}$ will be optimal if the firm is sufficiently large. Moreover, as discussed at the end of Section 4.3, Theorem 4 is not needed if we wish to focus on the cheapest contract to implement a given target action. The multiplicative production function is only necessary to implement the contract using stocks, i.e. (I2). With a multiplicative production function, the CEO's action affects the firm's return, and stocks are sensitive to the firm's return.

[20] One could argue that it is always possible to implement a contract with rebalancing and vesting, where the vesting fraction α_t and rebalancing target θ_t are complex functions of the past history, and so (I1) does not hinge on our assumptions (A1). However, such an implementation would be complex; the key role of assumptions (A1) is to allow θ_t and α_t to be history-independent.

A3. *Noise-before-action timing.* This timing assumption was convenient for the derivation of the contract by forcing the EF constraints to hold state-by-state. With reversed "action-before-noise" timing, the contract becomes complex even in a static model (see, e.g., Grossman and Hart (1983)). In particular, the solution typically does not feature a constant elasticity of pay to firm value. However, the paper's other insights, aside from (I2), remain valid. We sketch the general argument using a simple example.

Consider a one-period problem, in which principal minimizes the cost of providing incentives to exert effort $\overline{a}$, with log utility and "action-before-noise" timing. First, it can be shown that, with log utility, if $c(r)$ solves the problem when the agent's expected utility is U, then for any $z > 0$, $z \times c(r)$ solves the problem with expected utility $U + \ln z$. In other words, to deliver a higher expected utility, the principal must scale up all payments by the same fixed constant, regardless of the realized returns. The timing assumption only matters for the actual form of $c\,(r)$ (with "noise-before-action" timing, $c(r)$ has a particularly simple form: it is a multiple of e^{Cr} for some constant C; with "action-before-noise" timing, $c\,(r)$ a multiple of $e^{b(r)}$ for some general function b) – but the above "scaling" result holds regardless of the timing.

Moving to $T = L = 2$, the above claim means that the return at $t = 2$ affects pay at $t = 2$ multiplicatively. Therefore, the contract must have the form:

$$c_2(r_1, r_2) = e^{b_1(r_1)} e^{b_2(r_2)},$$

for $b_2 \equiv b$ and some function b_1. The PS constraint yields:

$$c_1(r_1) = e^{b_1(r_1) - k},$$

for the constant $k = \ln E[e^{-b_2(r_2)}]$, analogous to (11). Thus, $c_2(r_1, r_2)$ and $c_1(r_1)$ are affected by r_1 in the same manner. Finally, b_1 is the solution to a static problem where the CEO's utility of consumption is $2 \ln c$.

In sum, the two-period dynamic problem with private saving can be reduced to two static problems: solving for functions b_1 and b_2. Thus, while the static problem is complex, the dynamic model represents a simple extension: each static problem can be solved independently without complex history-dependence. Thus, much of (T) is preserved. Moreover, promised wealth at period 2 conditional on r_1 is multiplicative in $c_2(r_1, r_2)$ and so on, and so (I1) is preserved. At $t = 1$, the principal must invest the funds into an instrument that yields $e^{b_1(r_1)}$. At $t = 2$, regardless of r_1, she must invest the funds into an instrument that yields $e^{b_2(r_2)}$. With noise-before-action timing, $b_n\,(r) = \theta_n \times r$ so the instrument was a combination of cash and stock; with reversed timing, $b_n\,(r)$ is not linear in r and so in general the instrument will not be cash and stock, so we do not have (I2).

Appendix B shows that the contract retains the same form in continuous time, where the

noise and action are simultaneous.

5 Short-Termism

We now study how our basic contract changes when the agent can inflate the firm's returns, focusing on the log utility case for simplicity. Following on from Theorem, 4 we assume that $a_t^* = \overline{a} \ \forall \ t$. Short-termism is broadly defined to encompass any action that increases current returns at the expense of future returns. This includes real decisions such as scrapping positive-NPV investments (see, e.g., Stein (1988)) or taking negative-NPV projects that generate an immediate return but have a downside that may not manifest for several years (such as sub-prime lending), earnings management, and accounting manipulation.

We model short-termism in the following manner. In each period $t \leq L$, at the same time as taking action a_t, the agent also chooses a vector of myopic actions $m_t = \{m_{t,1}(r_{t+1}), ..., m_{t,M}(r_{t+M})\}$. A single myopic activity $m_{t,i}(r_{t+i}) \in [0, \overline{m}]$ (for an upper bound $\overline{m} > 0$) changes the returns from $r_s = a_s + \eta_s$ to

$$\begin{aligned} r'_t &= r_t + \lambda_i \left(E[m_{t,i}(r_{t+i})] \right) && \text{for } s = t, \\ r'_{t+i} &= r_{t+i} - m_{t,i}(r_{t+i}) && \text{for } s = t+i, \\ r'_s &= r_s && \text{for } s \neq t, t+i. \end{aligned}$$

Short-termism raises returns in period t by $\lambda_i \left(E[m_{t,i}(r_{t+i})] \right)$ (the function $\lambda_i (\cdot)$ will be specified shortly) and decreases them in period $t+i$ by $m_{t,i}(r_{t+i})$. This specification allows the CEO to engage in myopia state-by-state: the negative effect of short-termism $m_{t,i}$ depends on the realized return r_{t+i} and thus the state of nature η_{t+i}. Thus, the CEO can choose the states in which the costs of myopia are suffered. Giving the agent great freedom to inflate earnings restricts the set of admissible contracts that the principal can write to deter myopia, and thus leads to a simple solution to the contracting problem. This is similar to how specifying the noise before the action leads to tractability in the core model, as discussed in Section 2. In practice, CEOs can engage in short-termism by scrapping certain investments that pay off only in certain states of the world – for example, investing to increase the safety of a factory pays off if there is a disaster; expanding the capacity of a factory pays off only if demand turns out to be high.

We have $1 \leq i \leq M$, where i is the "release lag" of the myopic activity: the number of periods before its negative consequences become evident. For example, if the agent manipulates accounting to delay the realization of expenses for five years, $i = 5$. $M \leq \tau - L$ is the maximum possible release lag. The function $\lambda_i \left(E[m_{t,i}(r_{t+i})] \right)$ captures the efficiency of earnings inflation: a greater $\lambda_i (\cdot)$ means that a given future reduction in returns $E\left[m_{t,i}(r_{t+i})\right]$ translates into a

greater boost today. We assume $\lambda_i(0) = 0$, $\lambda_i' > 0$, $\lambda_i'' < 0$ and

$$q_i \equiv \lambda_i'(0) < \frac{e^{\underline{\eta} - M\overline{m}}}{E\left[e^{\eta}\right]}. \tag{31}$$

so that $0 < q_i < 1$. This assumption is sufficient to guarantee that all myopic actions are inefficient and create a first-order loss on firm value by reducing the expected terminal dividend, as proven in Appendix E.

5.1 Local Constraint

If the agent engages in a small myopic action $m_{t,i}(r_{t+i})$ at time t, his utility changes to the leading order by

$$E_t\left[\frac{\partial U}{\partial r_t}\right] q_i E_t\left[m_{t,i}\left(\widetilde{r}_{t+i}\right)\right] + E_t\left[-m_{t,i}(\widetilde{r}_{t+i}) E_t\left[\frac{\partial U}{\partial r_{t+i}} \mid \widetilde{r}_{t+i}\right]\right].$$

We require that, for every $m_{t,i}(r_{t+i}) \geq 0$, the change in utility is nonnegative, i.e.

$$E_t\left[\frac{\partial U}{\partial r_t}\right] q_i E_t\left[m_{t,i}\left(\widetilde{r}_{t+i}\right)\right] + E_t\left[-m_{t,i}(\widetilde{r}_{t+i}) E_t\left[\frac{\partial U}{\partial r_{t+i}} \mid \widetilde{r}_{t+i}\right]\right] \leq 0, \text{ i.e.}$$

$$E_t\left[\frac{\partial U}{\partial r_t}\right] q_i \int m_{t,i}\left(r_{t+i}\right) f(r_{t+i}) dr_{t+i} - \int m_{t,i}\left(r_{t+i}\right) f(r_{t+i}) E_t\left[\frac{\partial U}{\partial r_{t+i}} \mid \widetilde{r}_{t+i} = r_{t+i}\right] dr_{t+i} \leq 0,$$

This leads to the following No Myopia (NM) constraint:

$$\text{NM}: \forall r_{t+i}, E_t\left[\frac{\partial U}{\partial r_t}\right] q_i - E_t\left[\frac{\partial U}{\partial r_{t+i}} \mid \widetilde{r}_{t+i} = r_{t+i}\right] \leq 0. \tag{32}$$

To interpret the conditioning, consider the case $i = 3$. The second expectation is conditioned on $(r_s)_{s \leq t}$ and r_{t+3}, but not on r_{t+1} nor r_{t+2}.

5.2 The Contract

There are now three local constraints: EF, PS and NM. We seek the cheapest contract that satisfies these three constraints, i.e. induces zero myopia, zero private saving and high effort. The intuition behind implementing zero myopia is similar to that behind high effort as proven in Theorem 4: the benefits of preventing short-termism are multiplicative in firm size and thus orders of magnitude greater than the costs, which are a function of the CEO's salary. Relatedly, using a similar argument to Theorem 3, we conjecture that the contract that satisfies the three local constraints will also satisfy the global constraints if the function $\lambda_i(\cdot)$ (which captures the efficiency of inflation) is sufficiently concave, analogous to the sufficient condition on the convexity of the cost of effort $g(\cdot)$ in Theorem 3. Given the high complexity of the proofs of

Theorems 3 and 4, we do not provide analogous proofs here.

Proposition 2 below gives the cheapest contract that satisfies the three local constraints.

Proposition 2 *(Log utility, myopia possible.) The cheapest contract that satisfies the local constraints for high effort, zero private saving and zero myopia is as follows. In each period t, the CEO is paid c_t which satisfies:*

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

where θ_s and k_s are constants. The sensitivity θ_s is given by:

$$\theta_s = \begin{cases} \frac{\zeta_s}{1+\rho+...+\rho^{T-s}} \text{ for } s \leq L+M \\ 0 \text{ for } s > L+M \end{cases}, \tag{33}$$

with $\zeta_1 = g'(\bar{a})$. For $s > 1$, ζ_s is defined recursively as:

$$\zeta_s = \begin{cases} \max_{1 \leq i \leq M} \left\{ g'(\bar{a}), \frac{q_i}{\rho^i} \zeta_{s-i} \right\} \text{ for } s \leq L \\ \max_{s-L \leq i \leq M} \left\{ \frac{q_i}{\rho^i} \zeta_{s-i} \right\} \text{ for } L < s \leq L+M \end{cases}.$$

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

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

If private saving is possible, k_s is given by:

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

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

From (33), the possibility of short-termism has three effects on the contract sensitivity, which must change to prevent such actions. First, in the core model, there are two motivations for time-dependent vesting: consumption smoothing and the need to maintain sufficient equity in the DIA to satisfy the EF constraints in future periods. These motivations exist during the CEO's employment only and full vesting occurs in period L. Where myopia is possible, time-dependent vesting has an additional motivation – to satisfy the NM constraint in the current period, by preventing the CEO from inflating the current stock price and immediately cashing out. This motivation exists both during the CEO's employment and after retirement. Thus, gradual vesting continues after retirement and the account only fully vests in period $L+M$, since myopia allows the CEO to affect firm returns up to period $L+M$. While we are unaware of any large-scale studies, anecdotal evidence is consistent with such lock-ups. The severance

agreement of Stanley O'Neal (ex-CEO of Merrill Lynch) states that: "the unvested restricted stock and restricted stock units will continue to vest in accordance with their original schedules." During employment, equity grants are often restricted in practice: Kole (1997) finds a typical vesting horizon of 2-3 years. A number of firms are lengthening their horizons in the aftermath of the financial crisis: many commentators argued that short vesting periods in certain firms encouraged myopia in the crisis.[21]

Second, the contract sensitivity θ_t is higher in each period, because the contract must now satisfy NM as well as EF. Third, θ_t trends upwards more rapidly over time. Short-termism allows the CEO to increase the time-t return and thus his time-t consumption. Even though the return at time $t + i_t$ will be lower, the effect on the CEO's utility is discounted. Therefore, an increasing sensitivity is necessary to deter myopia, so that he loses more dollars in the future than he gains today to offset the effect of discounting. For example, in an infinite horizon model where myopia is impossible, (21) shows that the sensitivity is constant. (33) shows that the sensitivity is increasing over time if short-termism is possible.

The magnitude of the above three changes depends on the CEO's incentives to inflate earnings, which are determined by two forces. The benefit to the CEO of short-termism is that he boosts current returns and thus pay, which outweighs the negative effect on future returns owing to discounting. The discount rate ρ determines the size of this benefit. The cost is that myopia is inefficient, as the current boost to returns exceeds the future cost. For local myopic actions, the parameter q_i determines the size of the cost. Overall, when q_i is higher and ρ is lower, the CEO's incentives to inflate earnings are greater; thus, the CEO is given greater exposure to returns after retirement, and the contract sensitivity is higher in every period and increases more rapidly over time.

Moreover, all of the above changes to the sensitivity θ_t also affect the constant term k_t. Thus, if private saving is possible, the increase in θ_t causes the level of the contract to grow more rapidly over time, providing automatic saving for the agent. While the possibility of myopia only has a direct effect on the sensitivity of pay, this spills over into an indirect effect on the level of pay.

A specific example conveys the economics of the contract more clearly. Let $q_i = Q^i$ for some $Q \in (0, 1)$, i.e. a myopic action hidden for i periods increases current returns by Q^i, a factor that decreases at a constant rate Q per year of hiding. This natural benchmark allows for the slopes ζ_t in (33) to be defined explicitly rather than recursively. These are given as follows.

Corollary 2 *Suppose that* $Q \in (0, 1)$, $q_i = Q^i$. *If* $Q < \rho$, *then* $\zeta_t = g'(\overline{a})$ *for* $t \leq L$ *and*

[21] For example, Angelo Mozilo, the former CEO of Countrywide, sold over $100m of stock prior to his firm's collapse; Bebchuk, Cohen and Spamann (2010) estimate that top management at Bear Stearns and Lehman earned $1.4bn and $1bn respectively from cash bonuses and equity sales during 2000-8; a November 20, 2008 *Wall Street Journal* article entitled "Before the Bust, These CEOs Took Money Off the Table" provides further examples. Johnson, Ryan and Tian (2009) find a positive correlation between corporate fraud and unrestricted (i.e. immediately vesting) stock compensation.

$\zeta_t = g'(\overline{a})(Q/\rho)^{t-L}$ *for* $L < t \leq L+M$. *If* $Q \geq \rho$, *then* $\zeta_t = g'(\overline{a})(Q/\rho)^{t-1}$ *for* $t \leq L+M$.

We consider an infinite horizon model ($T = L = \infty$) for comparison with the sensitivity in the absence of myopia, $\theta_t = (1-\rho)g'(\overline{a})$ from (21). ζ_t depends on whether $Q \lessgtr \rho$, owing to the above trade-off arguments. If $Q < \rho$, myopia is sufficiently inefficient that the benefit is less than the cost. Thus, the contract in the core model (equation (21)) is already sufficient to deter short-termism and need not change. If $Q > \rho$, the CEO does have incentives to inflate earnings under the original contract, and so the sensitivity must increase to

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

The $(Q/\rho)^{t-1}$ term demonstrates that the sensitivity is not only greater in every period than in the core model, but is also increasing over time. The more impatient the CEO, the greater the incentives to inflate earnings, and so the greater the required increase in sensitivity over time to deter myopia. In a finite horizon model, θ_t is already increasing if myopia is impossible; the feasibility of short-termism causes it to rise even faster.

5.2.1 Numerical Example

We return to the last numerical example from Section 3.2.1 to demonstrate the effect of myopia on the contract. If $M = 1$, the contract changes from (23) to:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1 \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2 \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3 \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{q_1 r_4}{2} + \kappa_4 \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{q_1 r_4}{2} + \kappa_5.
\end{aligned}$$

The CEO's income now depends on r_4, otherwise he would have an incentive to boost r_3 at the expense of r_4. The sensitivity to r_4 depends on the efficiency of earnings inflation q_1; in the extreme, if $q_1 = 0$, myopia is impossible and so there is no need to expose the CEO to returns after retirement. 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 inflate earnings at $t = 1$ or $t = 2$ because there is no discounting, and so the negative effect of myopia on future returns reduces the CEO's lifetime utility by more than as the positive effect on current returns increases it. Appendix D allows for a variable cost of effort and shows that the possibility of short-termism forces the contract to change in $t \leq L$ even if there is no discounting.

6 Conclusion

This paper presents a new framework for studying CEO compensation in a fully dynamic model while retaining tractability. The model allows the CEO to consume in each period, privately save, and temporarily inflate returns. The model's closed-form solutions yield clear implications for the economic drivers of both the level of pay and the sensitivity of pay to performance. Pay depends on stock returns in the current and all past periods, and the sensitivity to a given return is constant over time. 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 short-termism is impossible, and rising if the horizon is finite or if short-termism is possible, even in the absence of career concerns. Deterring myopia also requires the CEO to remain sensitive to firm returns after retirement. By contrast, the feasibility of private saving only impacts the level of 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 mechanism that we call a "Dynamic Incentive Account". The CEO's expected pay is placed into an account, of which a certain proportion is invested in the firm's stock. The account features state-dependent rebalancing to ensure that, as the stock price changes, the CEO always has sufficient incentives to exert effort in the current period. It also features time-dependent vesting during employment, to ensure that the CEO exerts effort in future periods, and after retirement to deter myopia.

Our key results are robust to a broad range of settings: general CRRA utility functions, all noise distributions with interval support, and autocorrelated noise. However, our setup imposes some limitations, in particular that the CEO remains with the firm for a fixed period. Abstracting from imperfect commitment problems allows us to focus on a single source of market imperfection – moral hazard – and is common in the dynamic moral hazard literature (e.g. Lambert (1983), Rogerson (1985), Biais et al. (2007, 2009)). An interesting extension would be to allow for quits and firings. As is well-known (e.g. Bolton and Dewatripont (2005)), the possibility of quitting significantly complicates intertemporal risk-sharing since the agent may leave if his continuation wealth is low; firings may provide an additional source of incentives (as analyzed by DeMarzo and Sannikov (2006) and DeMarzo and Fishman (2007) in a risk-neutral model).[22] We leave those extensions to future research.

[22] The implementation of the contract via the DIA will involve the CEO forfeiting a portion of the account if he leaves early. Indeed, such forfeiture provisions are common in practice (see Dahiya and Yermack (2008)).

A Proofs

A.1 Proof of Theorem 1

This is a direct corollary of Theorem 2.

A.2 Proof of Proposition 1

The present value of future pay on the equilibrium path is given by $A_t = E_t\left[\sum_{s=t}^{T} e^{-R(s-t)}c_s\right]$, where $c_t = c_0 e^{\sum_{s=1}^{t}\theta_s r_s + k_s}$. We have $A_{t-1} - c_{t-1} = e^{-R}E_{t-1}[A_t]$. The contract in Theorem 1 implies $A_t = E_{t-1}[A_t]\, e^{\theta_t r_t}/E_{t-1}\left[e^{\theta_t r_t}\right]$. Thus,

$$A_t = (A_{t-1} - c_{t-1})\, e^R \frac{e^{\theta_t r_t}}{E_{t-1}\left[e^{\theta_t r_t}\right]}.$$

A_t is obtained by investing the residual value $A_{t-1} - c_{t-1}$ in a continuously rebalanced portfolio with a proportion θ_t in stock and the remainder in interest-bearing cash. ($1 invested at time $t-1$ in such an asset yields $e^R e^{\theta_t r_t}/E_{t-1}\left[e^{\theta_t r_t}\right]$, because both stock and cash have an expected return of R.) This is precisely the implementation via a DIA.

To derive the vesting fractions, we have

$$\begin{aligned}\alpha_t = c_t/A_t &= c_t/E_t\left[\sum_{s=t}^{T} e^{-R(s-t)}c_s\right] \\ &= 1/E_t\left[\sum_{s=t}^{T} e^{-R(s-t)} e^{\sum_{n=t+1}^{s}\theta_n\eta_n + k_n}\right]\end{aligned} \tag{34}$$

In certain benchmark cases these terms collapse into simple expressions:

(i) If private saving is impossible, the IEE gives us that inverse discounted marginal utility $\rho^{-t}e^{-Rt}c_t$ is a martingale. Thus $A_t = c_t\left(1-\rho^{T-t}\right)/(1-\rho)$ which yields $\alpha_t = c_t/A_t = (1-\rho)/\left(1-\rho^{T-t}\right)$.

(ii) If private saving is possible and the model horizon is infinite, the problem is stationary; given CRRA, the CEO consumes a constant fraction α of his wealth in each period and so $c_t = \alpha A_t$. We have:

$$\begin{aligned}k_s &= R + \ln\rho + \ln E\left[e^{-\theta(\bar{a}+\eta)}\right], \\ E\left[e^{\theta r_s + k_s}\right] &= E\left[e^{\theta\eta}\right] e^R \rho E\left[e^{-\theta\eta}\right] = e^R\rho_*,\end{aligned}$$

where

$$\rho_* = \rho E\left[e^{\theta\eta}\right] E\left[e^{-\theta\eta}\right].$$

Hence, for $s \geq t$,

$$E_t\left[e^{-R(s-t)}c_s\right] = c_t\rho_*^{s-t}$$

and

$$A_t = E_t\left[\sum_{s=t}^{\infty} e^{-R(s-t)}c_s ds\right] = E_t\left[\sum_{s=t}^{\infty} \rho_*^{s-t}c_t ds\right] = c_t/\left(1-\rho_*\right).$$

This yields $\alpha = c_t/A_t = 1 - \rho E\left[e^{\theta\eta}\right]E\left[e^{-\theta\eta}\right]$ as required.

A.3 Proof of Theorem 2

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 the IEE (if $\gamma = 1$) immediately give

$$\ln c_t(r_1, ..., r_t) = \ln c_L(r_1, ..., r_L) + \kappa'_t, \tag{35}$$

for some constants κ'_t independent of the returns.

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

$$\ln c_{t'}(r_1, ..., r_{t'}) = B(r_1, ..., r_t) + \theta_{t'}r_{t'} + \sum_{s=t+1}^{t'-1}(\theta_s - \phi\theta_{s+1})r_s + \kappa_{t'}, \tag{36}$$

for some function B, constants κ_t, and θ_s as in the Theorem. The PS constraint yields

$$c_t^{-\gamma} = e^R\frac{J_{t+1}}{J_t}E_t\left[c_{t+1}^{-\gamma}\right] = E_t\left[e^{-\gamma\theta_{t+1}r_{t+1}}\right]e^{-\gamma B(r_1,...,r_t)+R-\gamma\kappa_{t+1}+\ln J_{t+1}-\ln J_t}. \tag{37}$$

We therefore have[23]

$$\ln c_t = B(r_1, ..., r_t) + \phi\theta_{t+1}r_t + \kappa_t, \tag{38}$$

for the appropriate constant κ_t.

The EF constraint requires that in the case when $a_t^* \in (0, \overline{a})$

$$0 \in \arg\max_{\varepsilon} U(r_1, ..., r_{t-1}, a_t^* + \eta_t + \varepsilon). \tag{39}$$

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

$$J_t{c_t}^{1-\gamma}\phi\theta_{t+1} + \frac{d}{d\varepsilon}B\left(r_1, ...r_{t-1}, a_t^* + \eta_t + \varepsilon\right)\sum_{m=t}^{T}J_m E_t\left(c_m^{1-\gamma}\right) = J_t{c_t}^{1-\gamma}g'(a_t^*), \tag{40}$$

$$\frac{d}{d\varepsilon}B\left(r_1, ...r_{t-1}, a_t^* + \eta_t + \varepsilon\right) = \frac{J_t\left(g'(a_t^*) - \phi\theta_{t+1}\right)}{\sum_{m=t}^{T}J_m\prod_{n=t+1}^{m}E_t\left[e^{(1-\gamma)\left[\theta_n(\varepsilon_n + a_n^* - \phi a_{n-1}^*) + (\kappa_n - \kappa_{n-1})\right]}\right]} := \theta_t - \phi\theta_{t+1}.$$

[23] Equation (38) can also be derived from the IEE if $\gamma = 1$.

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

$$E_t\left[c_m^{1-\gamma}\right] = c_t^{1-\gamma} E_t\left[e^{(1-\gamma)\sum_{n=t+1}^{m}\left[\theta_n(\varepsilon_n + a_n^* - \phi a_{n-1}^*) + (\kappa_n - \kappa_{n-1})\right]}\right] = c_t^{1-\gamma} \prod_{n=t+1}^{m} E_t\left[e^{(1-\gamma)\left[\theta_n(\varepsilon_n + + a_n^* - \phi a_{n-1}^*) + (\kappa_n - \kappa_{n-1})\right]}\right]$$

In the case when $a_t^* = \overline{a}$ in an analogous way we get:

$$\frac{d}{d\varepsilon_-} B\left(r_1, ...r_{t-1}, \overline{a} + \eta_t + \varepsilon\right) \geq \frac{J_t\left(g'(\overline{a}) - \phi\theta_{t+1}\right)}{\sum_{m=t}^{T} J_m \prod_{n=t+1}^{m} E_t\left[e^{(1-\gamma)\left[\theta_n(\varepsilon_n + a_n^* - \phi a_{n-1}^*) + (\kappa_n - \kappa_{n-1})\right]}\right]} \tag{41}$$

We now show that (41) binds. First, (41) implies that for any $r' \geq r$ (see EG, Lemma 4)

$$B_t\left(r_1, ...r_{t-1}, r'\right) - B_t\left(r_1, ...r_{t-1}, r\right) \geq (\theta_t - \phi\theta_{t+1})(r' - r), \tag{42}$$

and it can be inductively shown that $0 \leq \theta_t - \phi\theta_{t+1} \leq g'(\overline{a})$. Consider now the contract $\left(c_s^0\right)_{s \leq T}$ that coincides with $\left(c_s\right)_{s \leq T}$ for $s < t$, and for $s \geq t$ c_s^0 are as in (36) and (38) with $B(r_1, ..., r_t) = B(r_1, ..., r_{t-1}) + (\theta_t - \phi\theta_{t+1})r_t$, where $B(r_1, ..., r_{t-1})$ is chosen to satisfy

$$E_{t-1}\left[\frac{\left(c_t^0\right)^{1-\gamma}\left(r_1, ..., r_t\right)}{1-\gamma}\right] = E_{t-1}\left[\frac{\left(c_t\right)^{1-\gamma}\left(r_1, ..., r_t\right)}{1-\gamma}\right]. \tag{43}$$

Condition (42) guarantees that the random variable $\ln c_t\left(r_1, ...r_{t-1}, \widetilde{r}_t\right)$ is weakly more dispersed than $\ln c_t^0\left(r_1, ...r_{L-1}, \widetilde{r}_L\right)$. It also follows from the EF that both $\ln c_t\left(r_1, ...r_{t-1}, \cdot\right)$ and $\ln c_t^0\left(r_1, ...r_{t-1}, \cdot\right)$ are weakly increasing. These facts, together with (43), imply that for the convex function ψ and increasing function ξ, where $\psi^{-1}(x) = \frac{x^{1-\gamma}}{1-\gamma}$, $\xi(x) = \frac{e^{(1-\gamma)x}}{1-\gamma}$ for $\gamma \neq 1$ and $\psi(x) = e^x$, $\xi(x) = x$ for $\gamma = 1$, we have (see EG, Lemmas 1 and 2):

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

In the same way we show that $E_{t-1}[c_s^0(r_1, ..., r_s)] \leq E_{t-1}[c_s(r_1, ..., r_s)]$ for any $s \geq t$. Consequently the contract $\left(c_s^0\right)_{s \leq T}$ is cheaper than $\left(c_s\right)_{s \leq T}$, and so indeed (41) must bind.

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. Writing $\kappa_0 = \ln c_0$ and $k_s = \kappa_s - \kappa_{s-1}$ establishes (24).

We now determine the values of the constants k_s. First, we have $c_0^{-\gamma} = e^{-\gamma \ln c_0} = e^{Rt} J_t E\left[c_t^{-\gamma}\right]$

for $t \leq T$ for all t. This yields, for all t:

$$\gamma \sum_{s=1}^{t} k_s = Rt + \ln J_t + \sum_{s=1}^{t} \ln E\left[e^{-\gamma\theta_s(\varepsilon_s + a_s^* - \phi a_{s-1}^*)}\right],$$

yielding (26). When the PS constraint is not imposed, we use (7) to derive (10) analogously.

Equation (25) becomes simpler in the limit case $L = T = \infty$ when $a_t^* = \overline{a}\ \forall t$. Then the problem is stationary, and θ and k are constant. To characterize them, define $f(\theta) = E\left[e^{(1-\gamma)[\theta(\epsilon+\bar{a}(1-\phi))+k]}\right]$ where $\gamma k = R + \ln\rho + \ln E\left[e^{-\gamma\theta(\epsilon+\bar{a}(1-\phi))}\right]$, so that

$$f(\theta) = E\left[e^{(1-\gamma)\theta\epsilon}\right]\left(E\left[e^{-\gamma\theta\epsilon}\right]\right)^{\frac{1-\gamma}{\gamma}} e^{\frac{1-\gamma}{\gamma}(R+\ln\rho)}.$$

Then from (25), we have $\theta = \frac{g'(\bar{a})-\phi\theta}{\sum_{s=t}^{\infty}[\rho f(\theta)]^{s-t}} + \phi\theta$, i.e.

$$\theta = (g'(\bar{a}) - \phi\theta)(1 - \rho f(\theta)) + \phi\theta. \tag{44}$$

In the limit of small time intervals, when $\phi = 0$, θ satisfies:

$$\begin{aligned}\theta &= g'(\bar{a})\left(-\ln\rho + \frac{\gamma-1}{\gamma}(R+\ln\rho) + \frac{\gamma-1}{2}\theta^2\sigma^2\right)\\ &= g'(\bar{a})\left(\frac{(\gamma-1)R - \ln\rho}{\gamma} + \frac{\gamma-1}{2}\theta^2\sigma^2\right)\end{aligned}$$

The value of θ is the root that goes to a finite limit as $\gamma \to 1$:

$$\theta = \frac{1 - \sqrt{1 - 2(\gamma-1)\sigma^2 g'(\bar{a})^2 \frac{(\gamma-1)R - \ln\rho}{\gamma}}}{(\gamma-1)\sigma^2 g'(\bar{a})} \tag{45}$$

Indeed, as $\gamma \to 1$, $\theta \to g'(\bar{a})(-\ln\rho)$, which is the solution from the log case in the limit of small time intervals.

A.4 Proof of Theorem 3

We divide the proof into the following steps.

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:

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

where $\beta = sign(1-\gamma)$, 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 so that $u(c,x)$ is jointly concave in both arguments.

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:

$$\max_{x_t,c_t \text{ adapted}} E\left[U\left((c_t)_{t\leq T},(x_t)_{t\leq L}\right)\right], \tag{46}$$

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

$$\ln y_t=\ln c_0+\sum_{s=1}^{t}\theta_s(\eta_s+f(x_s)-\phi(\eta_{s-1}+f(x_{s-1})))+\sum_{s=1}^{t}k_s, \tag{47}$$

for $f(x_s)=a_s^*$ for $s>L$. Problems (46) and (30) are equivalent: $(x_t)_{t\leq L}$ and $(c_t)_{t\leq T}$ solve (46) if and only if $(f(x_t))_{t\leq L}$ and $(c_t)_{t\leq T}$ solve (30). 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 $c_t^*(\eta)=c_0\exp\left(\sum_{s=1}^{t}\theta_s(\eta_s+f(x_s^*)-\phi(\eta_{s-1}+f(x_{s-1}^*)))+\sum_{s=1}^{t}k_s\right)$, the consumption for the recommended sequence of leisure on the path of noises $\eta=(\eta_t)_{t\leq T}$ (where $f(x_t^*)=a_t^*$), under no saving. For any path of noises $\eta=(\eta_t)_{t\leq T}$ we introduce the "upper linearization" utility function $\widehat{U}_\eta$:

$$\widehat{U}_\eta\left((c_t)_{t\leq T},(x_t)_{t\leq L}\right)=U+\sum_{t=1}^{T}(c_t-c_t^*(\eta))\frac{\partial U}{\partial c_t}+\sum_{t=1}^{L}(x_t-x_t^*)\frac{\partial U}{\partial x_t}, \tag{48}$$

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 $(c_t^*(\eta))_{t\leq T},(x_t^*)_{t\leq L})$. 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:

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

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

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

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 about when he consumes income y_t, and so we can evaluate $E\widehat{U}_\eta$ for $c_t\equiv y_t$. Since the agent has no motive to save, we only need to show that he has no motive to change leisure (and thus effort).[24] We also let utility be a function of $(x_t)_{t\leq L}$ since it

[24]For the same reason, it is satisfactory that we have linearized utility at the recommended consumption level.

fully determines 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* $\widetilde{U}_\eta\left((x_t)_{t\leq L}\right)=\widehat{U}_\eta\left((y_t)_{t\leq T},(x_t)_{t\leq L}\right)$ *for* $\widehat{U}_\eta$ *defined as in (48) and* y_t *as in (47), and consider the following maximization problem:*

$$\max_{x_t\ adapted} E\left[\widetilde{U}_\eta\left((x_t)_{t\leq L}\right)\right]. \tag{49}$$

If the target leisure level $(x_t^*)_{t\leq L}$ *solves the maximization problem (49) then* $(c_t^*)_{t\leq T}$ *and* $(x_t^*)_{t\leq L}$ *solve the maximization problem (46).*

Step 3. Pathwise concavity of utility in leisure for $\gamma=1$. We must demonstrate that expected utility is jointly concave in leisure $(x_t)_{t\leq L}$ if the cost function g is sufficiently convex. For $\gamma=1$, we can do so by proving pathwise concavity, i.e. that $\widehat{U}_\eta$ is concave for every path of noises. (We will deal with the case $\gamma\neq 1$ in step 4). We have:

$$\widetilde{U}_\eta\left((x_t)_{t\leq L}\right)=\sum_{t=1}^{T}\rho^t(\ln c_t^*(\eta)-1)+\sum_{t=1}^{L}\rho^t x_t+\sum_{t=1}^{T}e^{\sum_{s=1}^{t}\theta_s(f(x_s)-a_s^*-\phi(f(x_{s-1})-a_{s-1}^*))+t\ln\rho}. \tag{50}$$

Joint concavity of (50) in $(x_t)_{t\leq L}$ is equivalent to the joint concavity of "PV of income" function

$$I\left((x_t)_{t\leq L}\right)=\sum_{t=1}^{T}e^{\sum_{s=1}^{t}\theta_s(f(x_s)-a_s^*-\phi(f(x_{s-1})-a_{s-1}^*))+t\ln\rho}. \tag{51}$$

To prove the latter we will use the following general Lemma.

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

$$I((b_t)_{t\leq T})=\sum_{t=1}^{T}e^{\sum_{s=1}^{t}j_s(b_s)},$$

where $b_s\in\mathbb{R}$ *and all* j_s *are twice differentiable functions. Suppose that for every* s*:*

$$\sup\left[2Cj_s'^2+j_s''\right]\leq 0 \tag{52}$$

for $C=\sum_{n=0}^{T}e^{n\sup j_t/2}$*. Then the function* I *is concave.*

Loosely speaking, the Lemma states that, if j_s are sufficiently concave, 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 $(b_t)_{t\leq T}$. This is non-trivial to prove when $T\to\infty$: for sufficiently large t, $\exp\left(tj\left(b\right)\right)$

Since expected linearized utility does not depend on the agent's saving strategy, we can evaluate it with respect to an arbitrary savings strategy such as no saving (i.e. consuming the recommended amount).

is a convex function of b, because its second derivative is $\exp(tj(b))\, t\left(tj'(b)^2 + j''(b)\right)$, which is positive for sufficiently large t. It is discounting (expressed by $\rho < 1$) that allows the income function to be concave.

We use Lemma 2 to prove the following result.

Lemma 3 *(Concavity of present value of income.) The present value of income*

$$I\left((x_t)_{t\leq L}\right) = \sum_{t=1}^{T} e^{\sum_{s=1}^{t}\theta_s(f(x_s)-a_s^*-\phi(f(x_{s-1})-a_{s-1}^*))+t\ln\rho}$$

is jointly concave in leisure $(x_t)_{t\leq L}$.

Step 4. Concavity of expected utility in leisure for $\gamma \neq 1$. When $\gamma \neq 1$, linearized utility $\widetilde{U}_\eta$ is:

$$\begin{aligned}\widetilde{U}_\eta\left((x_t)_{t\leq L}\right) &= \sum_{t=1}^{L}\frac{\gamma}{1-\gamma}\rho^t c_t^*(\eta)^{1-\gamma}\left(\frac{x_t}{(\beta x_t^*)^{1-\gamma}}\right)\\ &+\sum_{t=1}^{T}\rho^t(\beta x_t^*)^\gamma c_0^{1-\gamma}e^{\sum_{s=1}^{t}\theta_s(f(x_s)-\gamma a_s^*-\phi(f(x_{s-1})-\gamma a_{s-1}^*)+(1-\gamma)\varepsilon_s)+(1-\gamma)k_s}. \qquad (53)\end{aligned}$$

Unlike when $\gamma = 1$, the second term in (53), i.e. the "PV of income function', now depends on noise η. We therefore cannot prove pathwise concavity of linearized utility, and instead prove concavity of expected utility directly.

Expected utility is given by

$$\begin{aligned}&E\left[\widetilde{U}_\eta((x_t)_{t\leq L})\right]\\ &= E\left[\sum_{t=1}^{L}A_t x_t + \sum_{t=1}^{T}M_t(\eta)e^{\sum_{s=1}^{t}\left[\theta_s(f(x_s)-\gamma a_s^*-\phi(f(x_{s-1})-\gamma a_{s-1}^*))+\ln E\left(e^{(1-\gamma)\theta_s\varepsilon_s}\right)+(1-\gamma)k_s\right]+t\ln\rho-(1-\gamma)g(a_t^*)}\right]\\ &= E\left[\sum_{t=1}^{L}A_t x_t + M_T(\eta)\sum_{t=1}^{T}e^{\sum_{s=1}^{t}\left[\theta_s(f(x_s)-\gamma a_s^*-\phi(f(x_{s-1})-\gamma a_{s-1}^*))+\ln E\left(e^{(1-\gamma)\theta_s\varepsilon_s}\right)+(1-\gamma)k_s\right]+t\ln\rho-(1-\gamma)g(a_t^*)}\right],\end{aligned}$$

where $M_t(\eta) = e^{\sum_{s=1}^{t}\left[(1-\gamma)\theta_s\varepsilon_s-\ln E\left(e^{(1-\gamma)\theta_s\varepsilon_s}\right)\right]+(1-\gamma)\ln c_0}$ is a martingale. The second equality follows from the law of iterated expectations and $M_t(\eta)$ being a martingale.

We use Lemma 2 to prove the following result.

Lemma 4 *(Concavity of modified present value of income.) The modified present value of income*

$$I'((x_t)_{t\leq L}) = \sum_{t=1}^{T}e^{\sum_{s=1}^{t}\left[\theta_s(f(x_s)-\gamma a_s^*-\phi(f(x_{s-1})-\gamma a_{s-1}^*))+\ln E\left(e^{(1-\gamma)\theta_s\varepsilon_s}\right)+(1-\gamma)k_s\right]+t\ln\rho-(1-\gamma)g(a_t^*)},$$

for $f(x_s) = a_s^*$ if $s > L$, is pathwise jointly concave in leisure $(x_t)_{t \leq L}$.

We now conclude the proof of the Theorem. From Theorem 2, $E\widetilde{U}_\eta$ satisfies the first-order conditions at $(x_t^*)_{t \leq L}$. From step 4, $E\widetilde{U}_\eta$ is also concave in $(x_t)_{t \leq L}$, and so the target leisure level $(x_t^*)_{t \leq L}$ solves the maximization problem (49). Therefore, from Lemma 1, $(c_t^*)_{t \leq T}$ and $(x_t^*)_{t \leq L}$ solve the maximization problem (46), establishing the result.

A.5 Proof of Theorem 4

This proof is in the Internet Appendix.

A.6 Proof of Proposition 2

We now impose the NM constraint. Proceeding inductively as in the proof of Theorem 2, we have a contract of the form:

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

with k_t as in the statement of Proposition 2, and θ_t deterministic lowest nonnegative values such that the EF and NM constraints are satisfied, i.e.:

$$EF : g'(\bar{a}) \leq \theta_t \left(1 + \rho + \ldots \rho^{T-t}\right) \text{ for } t \leq L, \tag{54}$$

$$NM : \theta_t \left(\rho^t + ... + \rho^T\right) q_i \leq \theta_{t+i} \left(\rho^{t+i} + ... + \rho^T\right), \text{ for } 0 \leq t \leq L,\ 0 \leq i \leq M. \tag{55}$$

Defining $\zeta_t = \theta_t \left(1 + \rho + \ldots \rho^{T-t}\right)$, this can be rewritten:

$$\begin{aligned} g'(\bar{a}) &\leq \zeta_t \text{ for } t \leq L \\ \zeta_t q_i &\leq \rho^i \zeta_{t+i} \text{ for } 0 \leq t \leq L,\ 0 \leq i \leq M \end{aligned}$$

This yields the values described in the Proposition.

References

[1] Abraham, Arpad, Sebastian Koehne and Nicola Pavoni (2010): "On the First-Order Approach in Principal-Agent Models with Hidden Borrowing and Lending." Working paper, European University Institute.

[2] Acharya, Viral, Kose John and Rangarajan Sundaram (2000): "On the Optimality of Resetting Executive Stock Options." *Journal of Financial Economics* 57, 65-101.

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

[4] Bebchuk, Lucian A., Alma Cohen and Holger Spamann (2009): "The Wages of Failure: Executive Compensation at Bear Stearns and Lehman 2000-2008." *Yale Journal on Regulation* 27, 257-282

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

[6] Bennardo, Alberto, Pierre-Andre Chiappori and Joon Song (2009): "Perks as Second Best Optimal Compensations." Working Paper, Columbia University

[7] Bennedsen, Morten, Francisco Perez-Gonzalez and Daniel Wolfenzon (2009): "Do CEOs Matter?" Working Paper, Copenhagen Business School

[8] Bettis, Carr, John Bizjak, Jeffrey Coles and Swaminathan Kalpathy (2010): "Stock and Option Grants with Performance-Based Vesting Provisions." *Review of Financial Studies* 23, 3849-3888.

[9] Bhagat, Sanjai and Roberta Romano (2009): "Reforming Executive Compensation: Focusing and Committing to the Long-Term." *Yale Journal on Regulation* 26, 359-372

[10] Bhattacharyya, Sugato and Jonathan B. Cohn (2009): "The Temporal Structure of Equity Compensation." Working Paper, University of Michigan

[11] Biais, Bruno, Thomas Mariotti, Guillaume Plantin and Jean-Charles Rochet (2007): "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Review of Economic Studies* 74, 345-390

[12] Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet and Stéphane Villeneuve (2010): "Large Risks, Limited Liability and Dynamic Moral Hazard." *Econometrica* 78, 73-118

[13] Bolton, Patrick and Mathias Dewatripont (2005): "Contract Theory." MIT Press, Cambridge

[14] Boschen, John F. and Kimberly J. 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

[15] Brisley, Neil (2006): "Executive Stock Options: Early Exercise Provisions and Risk-Taking Incentives." *Journal of Finance* 61, 2487-2509

[16] Chaigneau, Pierre (2009): "The Optimal Timing of Compensation with Managerial Short-Termism." Working paper, HEC Montreal

[17] Cole, Harold L. and Narayana R. Kocherlakota (2001): "Efficient Allocations with Hidden Income and Hidden Storage." *Review of Economic Studies* 68, 523-542.

[18] Core, John E. and David F. Larcker (2002): "Performance Consequences of Mandatory Increases in Executive Stock Ownership." *Journal of Financial Economics* 64, 317-340.

[19] Cremers, Martijn and Darius Palia (2010): "Tenure and CEO Pay." Working Paper, Yale University.

[20] Dahiya, Sandeep and David Yermack (2008): "You Can't Take It With You: Sunset Provisions for Equity Compensation When Managers Retire, Resign, or Die." *Journal of Corporate Finance* 14, 499-511.

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

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

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

[24] Dittmann, Ingolf and Ko-Chia Yu (2010): "How Important are Risk-Taking Incentives in Executive Compensation?" Working paper, Erasmus University Rotterdam.

[25] Dittmann, Ingolf, Ernst Maug and Oliver Spalt (2010): "Sticks or Carrots? Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 65, 2015-2050.

[26] Edmans, Alex and Xavier Gabaix (2011a): "Tractability in Incentive Contracting." Working Paper, University of Pennsylvania.

[27] Edmans, Alex and Xavier Gabaix (2011b): "The Effect of Risk on the CEO Market." *Review of Financial Studies*, forthcoming.

[28] Edmans, Alex, Xavier Gabaix and Augustin Landier (2009): "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies* 22, 4881-4917.

[29] Fahlenbrach, Rüdiger and René M. Stulz (2009): "Managerial Ownership Dynamics and Firm Value." *Journal of Financial Economics* 92, 342-361.

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

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

[32] Garrett, Daniel and Alessandro Pavan (2010): "Managerial Turnover." Working Paper, Northwestern University.

[33] 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.

[34] 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.

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

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

[37] He, Zhiguo (2009a): "Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion." *Review of Financial Studies* 22, 859-892.

[38] He, Zhiguo (2009b): "Dynamic Compensation Contracts With Private Savings." Working Paper, University of Chicago.

[39] Hermalin, Benjamin E. and Michael S. Weisbach (2009): "Information Disclosure and Corporate Governance." Working Paper, University of California at Berkeley.

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

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

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

[43] Kocherlakota, Narayana R. (2004): "Figuring Out the Impact of Hidden Savings on Optimal Unemployment Insurance." *Review of Economic Dynamics* 7, 541-554.

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

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

[46] Lambert, Richard (1983): "Long-Term Contracts and Moral Hazard." *Bell Journal of Economics* 14, 441-452

[47] Lazear, Edward (1979): "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87, 1261-1284

[48] 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.

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

[50] Prendergast, Canice (2002): "The Tenuous Trade-Off between Risk and Incentives." *Journal of Political Economy*, 110, 1071-102

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

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

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

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

[55] Yang, Jun (2009): "Timing of Effort and Reward: Three-Sided Moral Hazard in a Continuous-Time Model." *Management Science* 56, 1568–1583.

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

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