Dynamic Incentive Accounts*

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

May 18, 2010

Abstract

We study optimal executive compensation in a dynamic framework that incorporates many important features of the CEO job absent from a static setting. Shocks to firm value may weaken the incentive effects of securities over time (e.g. drive options out of the money). The CEO can undo the contract by privately saving, and temporarily manipulate the stock price. Despite the complex setup, we obtain a simple closed-form contract. It can be implemented by a “Dynamic Incentive Account”: the CEO’s expected pay is escrowed into an account, a fraction of which is invested in the firm’s stock and the remainder in cash. The account features state-dependent rebalancing and time-dependent vesting. 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, manipulation, 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 Gilles Chemla, Ingolf Dittmann, Phil Dybvig, Oliver Hart, Ken Feinberg, Mike Fishman, Zhiguo He, Eric Talley, and seminar participants at the AEA, Financial Research Association, Harvard Law School / Sloan Foundation Conference on Corporate Governance, Jackson Hole Finance Conference, LSE FMG Conference on Managers, Incentives and Organizational Structure, NBER, Washington University Conference on Corporate Finance, Arizona State, Caltech, Chicago, Colorado, Harvard, Northwestern, NYU, Toulouse and Wharton. Qi Liu and Andrei Savochkine provided excellent research assistance.
1 Introduction

Many classical models of CEO compensation consider only a single period, or multiple independent periods. However, the optimal static contract may be ineffective in a dynamic world. 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. By privately saving, the CEO can achieve a higher future income than intended by the contract, reducing his effort incentives. Short-term contracts can encourage the CEO to manipulate 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 the above complications, which are likely important features in reality. We take an optimal contracting approach which 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 manipulation is that the contract is typically very complex and can only be solved numerically, which makes it difficult to see the intuition and which features of the setting are driving which aspects of the contract. The main methodological contribution of the paper 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 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 manipulation, 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 only chooses effort. In the optimal contract, log pay is a linear function of current and all past stock returns. The rewards for exerting effort are spread over all future periods, to achieve intertemporal risk-sharing. The sensitivity to the return in a given period is constant over time – the first-period return has the same effect on first-period pay as it does on second-period pay. Moreover, in an infinite horizon model, this sensitivity is constant across periods – second-period pay is affected by first- and second-period returns to the same degree. With a finite horizon, the sensitivity is increasing over time, as found empirically by Gibbons and Murphy (1992). Pay is more sensitive to current than past returns, and the sensitivity to the current return intensifies as the CEO becomes older. This is for two reasons. First, holding constant the total lifetime monetary reward for effort (i.e. change in NPV of future pay, or change in wealth), this reward is spread across fewer periods and so the reward in the current period (i.e. change in pay) must increase. Second, the total lifetime monetary reward must also increase. As a risk-averse agent becomes older, a given increase in wealth provides less utility as he is forced to consume it over fewer periods; therefore, the reward must rise to maintain incentives. We thus generate a similar prediction to the model of Gibbons and Murphy, but without invoking career concerns.
While the contract slope 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. The prediction 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 depends on 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.

We finally allow the CEO to engage in manipulation, i.e. inflate the current stock price at the expense of future returns. In practice, this may entail changing accounting policies or scrapping positive-NPV projects. The optimal contract responds in two ways to prevent such behavior. The CEO’s income is now sensitive to firm returns after retirement, to deter him from inflating the stock price just before he leaves. 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 slope 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.

The optimal contract can be implemented in a 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 cash. Mathematically, the fraction of pay in stock equals the sensitivity of log pay to 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 work, which will increase firm value by 10%, or take a holiday that 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 halves, 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. While this is, to our knowledge, the first theory to advocate rebalancing in executive compensation, Core and Larcker (2002) and Fahlenbrach and Stulz (2009) find evidence that it occurs in reality. In addition, 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 also features gradual vesting: the CEO can only withdraw a percentage of the account in each period. This has two roles. First, it addresses the effort problem in future periods, by ensuring that the CEO has sufficient equity in the future to induce effort. This role exists even if manipulation is not feasible, and requires vesting to be gradual during the CEO’s employment. Second, it addresses the manipulation 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. 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, and a number of firms are lengthening their horizons in response to the financial crisis. Gradual vesting is a more effective solution to manipulation 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 manipulation 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; how that wealth was amassed over time does not matter.

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 manipulation. We provide a theoretical framework to underpin these verbal arguments (in particular, showing that gradual vesting is optimal even if manipulation is not feasible), and explicitly derive the vesting fraction in a number of cases. 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.

The DIA also differs from the above proposals as it involves rebalancing. Even if the vesting horizon could be extended at little cost, it only solves myopia and does not ensure continued effort incentives over time – even if the CEO must hold onto his options, they have little incentive effect if they are out of the money. Moreover, existing theories demonstrate costs of lengthening vesting horizons, and so the optimal horizon is short. In Peng and Roell (2009), long vesting periods increase the CEO’s risk by delaying the rebalancing of stock for cash; Bhattacharyya and Cohn (2008) and Brisley (2006) show that this increased risk deters the CEO from taking risky projects. Such costs arise because vesting and rebalancing are the same event in these models: the CEO can only sell stock when it vests, and so long vesting prevents timely rebalancing and risk reduction. The first two papers argue that short-vesting stock is optimal; 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, this requires state-dependent vesting. Indeed, recent empirical studies (e.g. Bettis et al. (2008)) document that performance-based (i.e. state-dependent) vesting is increasingly popular, where vesting is accelerated upon high returns.¹ 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 manipulation – high returns permit sales of equity, but critically the proceeds remain in the account in case the returns are subsequently reversed. The combination of time-dependent vesting and state-dependent rebalancing thus achieves a different result from state-dependent vesting – the two separate features achieve the two goals of deterring manipulation and maintaining effort incentives.

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) 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 manipulation. Lacker and Weinberg (1989), Goldman and Slezak (2006), Peng and Roell (2009) and Hermalin and Weisbach (2009) study manipulation in a static setting. To our knowledge, He (2009b) is the only other dynamic model featuring effort, manipulation 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

¹State-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.
using the framework of Edmans and Gabaix (2009) (“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 manipulation, 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 manipulate. 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. 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.\(^2\)
The first-order approach replaces the agent’s incentive constraints against complex multi-period
deviations with weaker local constraints (first-order conditions), with the hope that the solution
to the relaxed problem satisfies all constraints.\(^3\) 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

\(^2\) 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.

\(^3\) 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 (2009). See also Kocherlakota
(2004) for the analytical challenges of dynamic agency problems with private savings.
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 manipulation, Section 6 concludes and Appendix A contains proofs of theorems. In the Online Appendix, Appendix B gives the continuous-time version of the contract, Appendix C provides analysis supporting some of the comparative statics, Appendix D extends the model to a time-varying cost of effort, and Appendix E contains proofs of lemmas and other claims.

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 $D_\tau$ (“earnings”) in the final period $\tau$, given by

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

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. $\eta_t$ 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 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 requiring the Holmstrom and Milgrom assumptions by specifying that, in each period $t$, the agent privately observes $\eta_t$ 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 $\eta_t$) and thus tightly restricts the set of admissible contracts, leading to a simple solution to the principal’s problem. 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 $\eta$ and $\alpha$ are simultaneous.

After $\alpha$ 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.

$$r_t = a_t + \eta_t. \quad (2)$$

With a slight abuse of terminology, we call $r_t$ the firm’s “return”. By observing $S_t$, the principal learns $r_t$, but not its components $a_t$ and $\eta_t$. 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.

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. $c_t = y_t$.

---

$r_t$ 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. We later derive sufficient conditions under which the optimal contract implements $\bar{a} \forall t$. The stock price is thus:

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

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

A fully general contract can involve the income $y_t$ depending on messages sent by the agent regarding $\eta_t$. We later derive a sufficient condition under which the optimal contract implements a fixed action, $\bar{a}$, in every period. Hence, on the equilibrium path, there is a one to one correspondence between $r_t$ and $\eta_t$, 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.

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.
The agent’s per-period utility over consumption \( c_t \in [0, \infty) \) and effort \( a_t \) is given by
\[
u_c h(a_t),
\] (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.\(^7\) The agent discounts future utility at rate \( \rho \), so that his total discounted utility is given by:
\[
U = \sum_{t=1}^{T} \rho^t u(c_t h(a_t)).
\] (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 the wage increases.\(^8\) 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.\(^9\) 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)

\(^7\)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.

\(^8\)Bennardo, Chiappori and Song (2009) show that a multiplicative utility function can rationalize perks.

\(^9\)See Bennedsen, Perez-Gonzalez and Wolfenzon (2009) for empirical evidence that CEOs have the same percentage effect on firm value, regardless of firm size.
constraint is that the agent achieves his reservation utility of \( u \), i.e.

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

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

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

for all alternative effort strategies \( \{\hat{a}_t, t = 1, \ldots L \} \) and feasible consumption strategies \( \{\hat{c}_t, t = 1, \ldots T \} \). 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 start with a conjecture that the optimal contract involves binding local constraints and, if firm size \( X \) is sufficiently high, \( a_t = \bar{a} \ \forall \ t \). Following this conjecture we (i) derive the necessary local constraints that a candidate contract 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) derive a sufficient condition under which the candidate contract is optimal among all contracts, i.e. the optimal contract implements \( a_t = \bar{a} \ \forall \ t \) (Theorem 4 in Section 4.3).

3 Log Utility

3.1 Local Constraints

A candidate contract must satisfy two local constraints. The effort (EF) constraint ensures that the agent exerts \( a_t = \pi \). 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 \( \{y_t, t = 1, \ldots T\} \), a consumption strategy \( \{c_t, t = 1, \ldots T\} \) and
an effort strategy \( \{a_t, t = 1, \ldots, L\} \). Recall that \( y_t \) and \( c_t \) depend on the entire history \( (r_1, \ldots, r_t) \) and \( a_t \) depends on \( (r_1, \ldots, r_{t-1}, \eta_t) \). To capture history-dependence, \( E_t \) denotes the expectation conditional on \( (r_1, \ldots, r_t) \). We first address the EF constraint and consider a local deviation in the action \( a_t \) after history \( (r_1, \ldots, r_{t-1}, \eta_t) \). The effect on CEO utility is

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

Since \( \partial r_t/\partial a_t = 1 \) and \( \partial U/\partial a_t = \rho c_t h'(a_t)u'(c_t h(a_t)) \), the EF constraint is:

\[
\text{EF: } E_t \left[ \frac{\partial U}{\partial r_t} \right] \geq \rho c_t(-h'(a_t))u'(c_t h(a_t)). \tag{5}
\]

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

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

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 \( e^R \rho^t h(a_t)u'(c_t h(a_t)) \) 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; Section 4 considers $\gamma \neq 1$.

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

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

where $\theta_s$ and $k_s$ are constants. The slope $\theta_s$ is given by

$$\theta_s = \begin{cases} \frac{g'(\bar{a})}{1+\rho+\cdots+\rho^s} & \text{for } s \leq L, \\ 0 & \text{for } s > L. \end{cases}$$

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 $u$.

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

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

for some constants $\kappa_1$ (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.:

$$\ln c_1 = \theta_1 r_1 + \kappa_1, \quad \ln c_2 = \theta_2 r_1 + \theta_2 r_2 + \kappa_1 + k_2$$

for some constants $\theta_1, \theta_2, \kappa_1, k_2$. 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:

$$1 = E_1 \left[ \frac{c_1}{c_2} \right] = e^{(\theta_1 - \theta_{21}) r_1^1} E_1 \left[ e^{-\theta_2 r_2^2} \right]. \quad (14)$$

This must hold for all $r_1$. Therefore, $\theta_{21} = \theta_1$ and $k_2 = \ln E_1 \left[ e^{-\theta_2 r_2^2} \right]$, as in (11).

Next, consider total utility $U$:

$$U = \ln c_1 + \ln c_2 - g(a_1) - g(a_2) = 2\theta r_1 + \theta_2 r_2 - g(a_1) - g(a_2) + 2\kappa_1 + k_2.$$ 

From (5), the two EF conditions are

$$E_2 \left[ \frac{\partial U}{\partial r_1} \right] \geq g' (\bar{a}) \quad \text{and} \quad E_2 \left[ \frac{\partial U}{\partial r_2} \right] \geq g' (\bar{a}),$$

i.e.

$$2\theta \geq g' (\bar{a}), \quad \theta_2 \geq g' (\bar{a}).$$

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:

$$\frac{d (\ln c_2)}{dr_2} \geq g' (\bar{a}).$$

The cheapest contract involves this local EF condition binding, i.e.

$$d (\ln c_2) / dr_2 = g' (\bar{a}) \equiv \theta_2. \quad (15)$$

Integrating yields the contract:

$$\ln c_2 = \theta_2 r_2 + B (r_1), \quad (16)$$

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 $t = 1$:

$$1 = E_1 \left[ \frac{c_1}{c_2} \right] = E_1 \left[ \frac{c_1}{e^{\theta_2 r_2^2 + B(r_1)}} \right] = E_1 \left[ e^{-\theta_2 r_2^2} c_1 e^{-B(r_1)} \right]. \quad (17)$$

Hence, we obtain

$$\ln c_1 = B (r_1) + K, \quad (18)$$

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:

$$U = \ln c_1 + \ln c_2 + K' = \theta_2 r_2 + 2B(r_1) + 2K + K'. \quad (19)$$
We next apply (5) to (19) to yield: 
\[ 2B'(r_1) \geq g'(\pi) \].
Again, the cheapest contract involves this condition binding, i.e. 
\[ 2B'(r_1) = g'(\pi) \].
Integrating yields:
\[ B(r_1) = g'(\pi) \frac{r_1}{2} + K'' \].
Combining (20) with (18) yields: 
\[ \ln c_1 = g'(\pi) \frac{r_1}{2} + \kappa_1 \], for another constant \( \kappa_1 \). Combining (20) with (16) yields:
\[ \ln c_2 = g'(\pi) \left( \frac{r_1}{2} + r_2 \right) + \kappa_1 + k_2, \]
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. (9) shows that, in an infinite horizon model \( (T = \tau \to \infty) \), the sensitivity is constant and given by:
\[ \theta_t = \theta = (1 - \rho) g'(\pi). \]
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.

However, for any model with finite life \( T \) \( (9) \) shows that \( \theta_t \) is increasing over time. To understand the intuition for this increasing sensitivity, we distinguish between the increase in lifetime utility for exerting effort \( (\partial U/\partial a_t) \), the increase in current utility \( (\partial u_t/\partial a_t) \), the increase in current pay \( (\partial c_t/\partial a_t) \) and the increase in wealth \( (\partial A_t/\partial a_t) \), where \( A_t = E_t \left[ \sum_{s=t}^{T} e^{-R(s-t)} c_s \right] \) is wealth, i.e. the NPV of all future pay). 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; this in turn requires a greater increase in current pay \( (\partial c_t/\partial a_t) \). In addition, a constant lifetime increase in utility \( \partial U/\partial a_t \) requires the increase in wealth \( \partial A_t/\partial a_t \) to rise over time.\(^\text{10}\) As the CEO becomes older, a given increase in wealth provides him with less lifetime utility, because he is forced to consume it over fewer periods; since he is risk-averse, this is

\(^{10}\)In Theorem 1 we have \( A_t = c_t/a_t \) for some constant \( a_t \). Thus,
\[ \ln A_t = \ln A_0 + \sum_{s=1}^{t} \theta_s r_s + \sum_{s=1}^{t} k_s - \ln a_t + \ln a_0, \]
and so the sensitivity of \( \ln A_t \) to current returns is \( \theta_t \), which is increasing over time.
unattractive to him. Thus, to keep the lifetime increase in utility constant, a greater increase in wealth is required. By contrast, Gibbons and Murphy (1992) generate an increasing current sensitivity ($\partial c_t/\partial a_t$) because the lifetime increase in utility $\partial U/\partial a_t$ rises over time, to offset falling career concerns. In Garrett and Pavan (2009), $\partial c_t/\partial a_t$ 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 $\partial c_t/\partial a_t$ stems from the reduction in consumption smoothing possibilities as the CEO approaches retirement.

While $\theta_t$ 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): $\ln E[c_t/c_{t-1}] = k_t + \ln E[e^{\theta r_t}]$.

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

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

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:

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

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

$$\ln E[c_t/c_{t-1}] = R + \ln \rho + \theta^2 \sigma^2.$$

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. The model thus predicts a positive relationship between the wage and tenure, which is consistent with the common practice of seniority-based pay. 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 $\theta$, and the volatility of firm returns $\sigma$. 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

---

11 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.
him to do so himself. Furthermore, in a finite-horizon model, \( \theta_t \) 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.

**Proposition 1 (Cost of Private Savings).** Define \( \Lambda = (\text{Expected cost of contract imposing PS}) / (\text{Expected cost of contract without imposing PS}) \), and consider \( L = T = \infty \). We have \( \Lambda \geq 1 \) and:

\[
\Lambda = \frac{1 - \rho}{1 - \rho e^{\sigma^2 \tau^2} e^{-\frac{\rho^2 \sigma^2}{1 - \rho}}},
\]

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

The ratio \( \Lambda \) 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 \( \Lambda \) is closer to one when the agent is more patient.

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 \( \gamma \) as for log utility, we delay them until Section 4.

### 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 \( T = 3 \), \( L = 3 \), \( \rho = 0 \) and \( g' (\bar{a}) = 1 \). From (9), the contract is:

\[
\ln c_1 = \frac{r_1}{3} + \kappa_1
\]

\[
\ln c_2 = \frac{r_1}{3} + \frac{r_2}{2} + \kappa_2
\]

\[
\ln c_3 = \frac{r_1}{3} + \frac{r_2}{2} + \frac{r_3}{1} + \kappa_3
\]

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 1 in all periods, and the sensitivity in terms of cash compensation is \( \partial c_1 / \partial a_1 = \frac{1}{3} c_1 \), \( \partial c_2 / \partial a_2 = \frac{1}{2} c_2 \), \( \partial c_3 / \partial a_3 = c_3 \) and increasing over time.\(^\text{12}\) The sensitivity in terms of wealth is \( \partial A_1 / \partial a_1 = \frac{1}{3} (c_1 + c_2 + c_3) \), \( \partial A_2 / \partial a_2 = \frac{1}{2} (c_2 + c_3) \), \( \partial A_3 / \partial a_3 = c_3 \) and also increasing.

\(^{12}\)Since \( \rho = 0 \), the growth rate of consumption is non-negative, so \( c_3 > c_2 > c_1 \).
We now consider $T = 5$, so that the CEO lives after retirement. The contract is now:

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

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

From Theorem 1, we have

\[
\ln c_t - \ln c_{t-1} = \theta_t r_t + k_t.
\]

The percentage change in CEO pay is linear in the firm’s return $r_t$, i.e. the percentage change in firm value. The relevant measure of incentives is thus the elasticity of CEO pay to firm value; this elasticity must be $\theta_t$ for incentive compatibility and is independent of firm size. Empiricists have used a number of 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)).

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.

In real variables, percent-percent incentives equal the fraction of total pay that is comprised of stock. Thus, the contract can be implemented in the following simple manner. The present value of the CEO’s expected pay is escrowed into a “Dynamic Incentive Account” (“DIA”) at

---

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.
the start of \( t = 1 \). A proportion \( \theta_1 \) is invested in the firm’s stock and the remainder in cash. At the start of each subsequent period \( t \), the DIA is rebalanced so that the proportion invested in the firm’s stock is \( \theta_t \).\(^{14}\) 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 \( \theta_t \). 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 \( \theta_t \). 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 maintain a constant multiple) and implies targeting of percent-percent incentives.

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 \( \alpha_t \)

\(^{14}\)The justification is as follows. Consider the account value \( A_t = E_t \left[ \sum_{s=t}^{T} e^{-R(s-t)} c_s \right] \). We have \( A_{t-1} - c_{t-1} = e^{-R} E_{t-1} \left[ A_t \right] \). The contract in Theorem 1 implies \( A_t = E_{t-1} \left[ A_t \right] e^{\theta_t r_t} / E_{t-1} \left[ e^{\theta_t r_t} \right] \). Thus, \( A_t = (A_{t-1} - c_{t-1}) e^{\theta_t r_t} E_{t-1} \left[ e^{\theta_t r_t} \right] \).

\(^{15}\)Acharya, 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 \( \theta_t \). Thus, ownership must fall (rise) with good (bad) performance to keep \( \theta_t \) constant.
in each period for consumption. The fraction \( \alpha_t \) is history-independent\(^{17} \); we will later derive \( \alpha_t \) in specific cases. This gradual vesting is to ensure 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. 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 shows that gradual vesting is optimal even if manipulation is not possible.

We calculate the vesting percentage in a number of benchmark cases. Recall

\[
A_t = E_t \left[ \sum_{s=t}^{T} e^{-R(s-t)}c_s \right]
\]

\(^{(24)}\)

denotes the value of the DIA at date \( t \), i.e. the present value of future pay on the equilibrium path, where \( c_t = c_0 e^{\theta \tau_T + k_{s,t}} \). While \( A_t \) typically involves a complex sum of very many terms, in certain benchmark cases these terms collapse into simple expressions. If private saving is impossible, the IEE gives us that inverse discounted marginal utility \( \rho^{-t}e^{-Rt}c_t \) is a martingale. Thus \( A_t = c_t (1 - \rho^{T-t}) / (1 - \rho) \) and the vesting fraction is \( \alpha_t = c_t / A_t = - (1 - \rho) / (1 - \rho^{T-t}) \).

In an infinite horizon model, the vesting fraction is \( \alpha = 1 - \rho \) and time-independent, just like the contract sensitivity \( \theta_t \). If the horizon is finite, \( \alpha_t \) 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.

We return to an infinite horizon model but now allow for private saving. Since the problem is stationary and the CEO exhibits CRRA, he consumes a constant fraction \( \alpha \) of his wealth in each period and so \( c_t = \alpha A_t \). If noise \( \eta_s \) is i.i.d., we find \( \alpha = 1 - \rho E \left[ e^{\theta \eta} \right] E \left[ e^{-\theta \eta} \right] < 1 - \rho \).\(^{18} \)

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.

Note that the DIA 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:

\(^{17} \)We have \( \alpha_t = c_t / E_t \left[ \sum_{s=t}^{T} e^{-R(s-t)}c_s \right] = 1 / E_t \left[ \sum_{s=t}^{T} e^{-R(s-t)}e^{\sum_{n=1}^{\tau_T} \theta \eta_n + k_{n,s}} \right] \).

\(^{18} \)We have \( k_{s,t} = R + \ln \rho + \ln E \left[ e^{\theta(\bar{\eta} + s)} \right] \) and \( E \left[ e^{\theta r_{s,t} + k_s} \right] = E \left[ e^{\theta \eta} \right] e^{R \rho E \left[ e^{-\theta \eta} \right]} = e^{R \rho s} \), where \( \rho_s = \rho E \left[ e^{\theta \eta} \right] E \left[ e^{-\theta \eta} \right] \). Hence, for \( s \geq t \), \( E_t \left[ e^{-R(s-t)}c_s \right] = c_t \rho_{s-t} \) and \( A_t = E_t \left[ \sum_{s=t}^{\infty} e^{-R(s-t)}c_s ds \right] = E_t \left[ \sum_{s=t}^{\infty} \rho_{s-t} c_t ds \right] = c_t / (1 - \rho_s) \).
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 is a technical section that may be skipped in a first reading. 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 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 \( \eta_t \) are uncorrelated. However, in private firms, there is no stock return, and so alternative signals of effort or trust are 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 \( \eta_t \) follows an AR(1) process with autoregressive parameter \( \phi \), i.e. \( \eta_t = \phi \eta_{t-1} + \varepsilon_t \), \( \phi \in [0, 1] \), where \( \varepsilon_t \) are independent and bounded above and below by \( \underline{\varepsilon} \) and \( \bar{\varepsilon} \).

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 \( J_t = \rho^t e^{-(1-\gamma)\eta(t)} \) for \( t \leq L \) and \( J_t = \rho^t \) otherwise.

Theorem 2 (General CRRA utility, autocorrelated noise, with the PS constraint.) The cheapest contract that satisfies the local constraints and implements \( a_t = \bar{a} \) \( \forall t \) is as follows. In each period \( t \), the CEO is paid \( c_t \) which satisfies:

\[
\ln c_t = \ln c_0 + \sum_{s=1}^{t} \theta_s (r_s - \phi r_{s-1}) + \sum_{s=1}^{t} k_s, \tag{25}
\]

where \( \theta_s \) and \( k_s \) are constants and \( r_0 = 0 \). The slope \( \theta_s \) is given by:

\[
\theta_t = \begin{cases} 
\frac{J_t (\rho^t \eta(t) - \phi \theta_{t+1})}{\sum_{j=t}^{L} J_j \prod_{n=t+1}^{L} E_t [e^{(1-\gamma)\eta_n (e^{(1-\gamma)\theta_n + \pi(1-\gamma)}) + k_n}] + \phi \theta_{t+1}} & \text{for } t \leq L, \\
0 & \text{for } t > L.
\end{cases} \tag{26}
\]
The constant $k_t$ is given by:

$$\gamma k_t = R + \ln \rho - (1 - \gamma)g(\pi)1_{t=L+1} + \ln E \left[ e^{-\gamma \theta_t (\tau_t + \pi(1-\phi))} \right] \text{ for } t \leq T. \quad (27)$$

The initial condition $c_0$ is chosen to give the agent his reservation utility $u$. If $L = T = \infty$, the slope (29) simplifies to a constant $\theta_t = \theta$, where $\theta$ is given by (43) in Appendix A. In the limit of small time intervals, and when $\phi = 0$, we have:

$$\theta = \frac{1 - \sqrt{1 - 2(\gamma - 1)\sigma^2 g'(\bar{\alpha})^2 (\gamma - 1)\ln \rho}}{(\gamma - 1)\sigma^2 g'(\bar{\alpha})}. \quad (28)$$

and $k_t = k = (R + \ln \rho)/\gamma - \theta \bar{\alpha} - \gamma \theta^2 \sigma^2/2$.

Equation (25) 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, $\gamma$ only affects the specific values of $\theta$ and $k$ rather than the functional form. The time trend of the contract slope and the implementation via the DIA remain the same. The difference is that the parameters $\theta$ and $k$ are somewhat more complex. To understand the economic forces that determine $\theta$, consider the benchmark case where $\phi = 0$ and $L = T$. Then, the slope (25) becomes

$$\theta_t = \frac{\sum_{s=t}^{T} E \left[ J_s c_s^{1-\gamma} \right] g'(\pi)}{\sum_{s=t}^{T} E \left[ J_s c_s^{1-\gamma} \right] g'(\pi)}, \quad (29)$$

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 slope increases over time up to $\theta_T = g'(\pi)$ and is steeper if the agent is more risk averse (higher $\gamma$) and less patient (lower $\rho$), 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 (29). Since future rewards are insufficient to induce effort, the CEO must be given a higher sensitivity to current consumption.

Equation (25) shows that, with autocorrelated signals, the optimal contract links the percentage change in CEO pay in period $t$ to innovations in the signal $(\tau_t - \phi \tau_{t-1})$ 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.$^{19}$

$^{19}$Similarly, if there is an industry-wide component to $\tau_t$, the optimal contract will filter out this component, just as it filters out $\phi \tau_{t-1}$. 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:

\[
\ln y_t = \ln c_0 + \sum_{s=1}^{t} \theta_s (a_s + \eta_s - \phi(a_{s-1} + \eta_{s-1})) + \sum_{s=1}^{t} k_s, \tag{30}
\]

The following Theorem states that if the cost function \( g \) is sufficiently convex, the CEO has no profitable global deviation.

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

\[
\max_{a_t, c_t \text{ adapted}} E \left[ \sum_{t=1}^{T} \rho^t u \left( c_t e^{-g(a_t)} \right) \right] \tag{31}
\]

with \( \sum_{t=1}^{T} e^{-r_t} (y_t - c_t) \geq 0 \) and \( y_t \) satisfying (30). If function \( g \) is sufficiently convex, i.e. \( \inf_a g''(a) \) is sufficiently large, the solution of this problem is \( c_t \equiv y_t, t \leq T \), and \( a_t \equiv \tau, t \leq L \). There is no global deviation from the recommended policy that makes the agent better off.

The proof, in the Appendix, may be of general methodological interest. It involves three 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,\(^{20}\) 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.

\(^{20}\)See Dittmann and Yu (2009) 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.
4.3 The Optimality of Maximum Effort

This section derives conditions under which the principal wishes to implement the highest effort level \( a_t = \bar{a} \) in every period and after every history.

**Theorem 4** (Maximum effort is optimal.) Assume that 
\[ \inf_{\eta \in (\underline{\eta}, \overline{\eta})} f(\eta) > 0 \text{ and } \sup_{a \in (\underline{a}, \overline{a})} g''(a) / g'^2(a) < \infty, \]
where \( f \) is the probability density of \( \eta \). There exists \( X_* \) such that if baseline firm size \( X > X_* \), implementing \( a_t = \bar{a} \) as in Theorems 1 and 2 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 maximum effort contract. The benefits are multiplicative in firm size. The costs comprise the direct disutility from working (which are multiplicative in the CEO’s wage), the risk premium required to compensate the CEO for a variable contract, and the change in CEO’s informational rent (which are both also a function of the CEO’s wage). Since the CEO’s wage is substantially smaller than firm size, the benefits of maximum effort outweigh the costs. In practice, a maximum 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 CEO can work while remaining productive. Under the interpretation of \( a \) as rent extraction, \( \pi \) 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 \( \eta \)) 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 private 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 maximum 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 \( \theta_t \) 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 manipulation is possible.

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

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 \( \theta_L \), irrespective of the level of pay in period \( L \) and independent of the history of past shocks. 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 \( \theta_L \).

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 \( c_2(r_1, r_2) \), 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

\[^{21}\text{We require } c_2 (r_1, r_2) = c_1 (r_1) f (r_2), c_3(r_1, r_2, r_3) = c_2 (r_1, r_2) g (r_2) \text{ etc., i.e. multiplicative separability.}\]
$C$ is a constant. Thus, rewards for performance are smoothed in a simple manner, and this smoothing is also independent of the history of past shocks — for example, the effect of $r_2$ on $c_2, \ldots, c_T$ is independent of $r_1$.\textsuperscript{22} Together, both assumptions mean that, although consumption is history-dependent, $\theta$ 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 $\alpha_1$, 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 $\theta_t$ 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 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 — how that wealth was affected by past returns is irrelevant. Since consumption depends on current wealth alone, the vesting fraction $\alpha_t$ is history-independent.\textsuperscript{23}

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. (II). 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.

\textsuperscript{22}With multiplicative preferences but without CRRA, the smoothing is complex and history-dependent. Consider a 2-period model with $u(c, a) = e^{c\beta}$. We have $c_2(r_1, r_2) = B(r_1) e^{\theta_2 r_2}$, and PS yields $e^{c_1 h(\varphi)} = E_1 \left[ e^{B(r_1) e^{\theta_2 r_2} h(\varphi)} \right]$. Even though $r_1$ has a multiplicative effect on $c_2$, solving for the magnitude of this effect $B(r_1)$ is highly complex.

\textsuperscript{23}One could argue that one could always implement a contract with rebalancing and vesting, where the vesting fraction $\alpha_t$ and rebalancing target $\theta_t$ are complex functions of the past history, and so (II) does not hinge on our assumptions (A1). However, such an implementation would be complex; the key role of assumptions (A1) is to allow $\theta_t$ and $\alpha_t$ to be history-independent.
Multiplicative preferences are not necessary for (E). In any model with manipulation, the CEO must remain tied to firm returns after he retires. The time trend in $\theta$ 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 $a_t = \varpi$ 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, $a_t = \varpi$ will be optimal if the firm is sufficiently large. Moreover, the contract does not require $a_t = \bar{a}$ to be optimal; if the principal wishes to implement a constant target action (as in, e.g., Dittmann and Maug (2007)), all the results hold. 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.

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 with log utility and reversed timing. It can be easily shown that 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 $U + \frac{z}{1-\rho} \ln z$. With “noise-before action” timing, $c(r)$ is a multiple of $e^{Cr}$ for some constant $C$; with reversed timing it is a multiple of $e^{b_1(r)}$ for some function $b_1$. Moving to $T = L = 2$, the above claim means that the contract must have the form:

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

for some function $b_2$. The PS constraint yields:

$$c_1(r_1) = e^{b_2(r_1) - k},$$

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

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 (II) is preserved. At \( t = 1 \), the principal must invest the funds into an instrument that yields \( e^{b(r_1)} \). At \( t = 2 \), regardless of \( r_1 \), she must invest the funds into an instrument that yields \( e^{b(r_2)} \). 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 Manipulation

We now study how our basic contract changes when the agent can manipulate the firm’s returns, focusing on the log utility case for simplicity. Manipulation is broadly defined to encompass any action that increases current returns at the expense of future returns. In addition to the literal interpretation of changing accounting policies, it can also involve 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).

We model manipulation 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 manipulations \( m_t = \{ m_{t,1}(r_{t+1}), \ldots, m_{t,M}(r_{t+M}) \} \). A single manipulative 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

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

Manipulation raises returns in period \( t \) by \( \lambda_i (E[m_{t,i}(r_{t+i})]) \) (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 manipulation state-by-state: the negative effect of manipulation \( m_{t,i} \) depends on the realized return \( r_{t+i} \) and thus the state of nature \( n_{t+i} \). Thus, the CEO can choose the states in which the costs of manipulation are suffered. Giving the agent great freedom to manipulate restricts the set of admissible contracts that the principal can write to deter manipulation, 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 manipulation 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 manipulation: the number of periods before the manipulation becomes evident. For example, if manipulation delays the realization of expenses for five years, $i = 5$. $M \leq \tau - L$ is the maximum possible release lag. The function $\lambda_i(E[m_{t,i}(r_{t+i})])$ captures the efficiency of manipulation: a greater $\lambda_i(\cdot)$ means that a given future reduction in returns $E[m_{t,i}(r_{t+i})]$ translates into a greater boost today. We assume $\lambda_i(0) = 0$, $\lambda_i' > 0$, $\lambda_i'' < 0$ and

$$q_i \equiv \lambda_i'(0) < \frac{e^{\phi-M\phi\tau}}{E[\epsilon^K]}.$$  

So that $0 < q_i < 1$. This assumption is sufficient to guarantee that all manipulations 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 manipulation $m_{t,i}(r_{t+i})$ at time $t$, his utility changes to the leading order by

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

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

$$E_t \left[ \frac{\partial U}{\partial r_t} \right] q_i E_t [m_{t,i}(\bar{r}_{t+i})] + E_t \left[ -m_{t,i}(\bar{r}_{t+i})E_t \left[ \frac{\partial U}{\partial r_{t+i}} | \bar{r}_{t+i} \right] \right] \leq 0,$$

i.e.

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

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

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

(33)

To interpret the conditioning, take the case $i = 3$. The second expectation is conditioned on $(r_s)_{s \leq t}$ and $r_{t+3}$, but not on the randomness $r_{t+1}$ and $r_{t+2}$.

We see that our concept of manipulation gives rise to a rather simple to use local condition, (33).

### 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 manipulation, zero private saving and max-
imum effort. The intuition behind implementing zero manipulation is similar to that behind maximum effort as proven in Theorem 4: the benefits of preventing manipulation 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 manipulation) 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, manipulation possible.) The cheapest contract that satisfies the local constraints for maximum effort, zero private saving and zero manipulation 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 \( \theta_s \) and \( k_s \) are constants. The slope \( \theta_s \) is given by:

\[
\theta_s = \begin{cases} 
\frac{\zeta_s}{1+\rho+\ldots+\rho^t} & \text{for } s \leq L + M \\
0 & \text{for } s > L + M 
\end{cases},
\]  

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

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

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

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

If private saving is possible, \( k_s \) is given by:

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

The initial condition \( c_0 \) is chosen to give the agent his reservation utility \( v \).

From (34), the possibility of manipulation has three effects on the optimal contract, which must change to prevent such actions. First, in the core model, there is a single motivation for time-dependent vesting: the need to maintain sufficient equity in the DIA to satisfy the EF constraints in future periods. This motivation exists during the CEO's employment only and
full vesting occurs in period $L$. Where manipulation 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 manipulation allows the CEO to affect firm returns up to period $L+M$. Many commentators have indeed argued that short vesting periods induced myopia in the recent financial crisis.\textsuperscript{24}

Second, the contract sensitivity $\theta_t$ is higher in each period, because the contract must now satisfy NM as well as EF. Third, $\theta_t$ trends upwards more rapidly over time. Manipulation 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 manipulation, 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 manipulation is impossible, (21) shows that the sensitivity is constant. (34) shows that the sensitivity is increasing over time if manipulation is possible.

The magnitude of the above three changes depends on the CEO’s incentives to manipulate, which are determined by two forces. The benefit to the CEO of manipulating is that he boosts current returns and thus pay, which outweighs the negative effect on future returns owing to discounting. The discount rate $\rho$ determines the size of this benefit. The cost is that manipulation is inefficient, as the current boost to returns exceeds the future cost. For local manipulations, the parameter $q_i$ determines the size of the cost. Overall, when $q_i$ is higher and $\rho$ is lower, the CEO’s incentives to manipulate are greater; thus, the CEO is given greater exposure to returns after retirement, and the contract slope is higher in every period and increases more rapidly over time.

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 manipulation 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 $\zeta_t$ in (34) to be defined explicitly rather than recursively. These are given as follows.

**Corollary 1** Suppose that $Q \in (0,1)$, $q_i = Q^i$. If $Q < \rho$, then $\zeta_t = g'(\pi)\frac{Q^i}{Q^t}$ for $t \leq L$ and $\zeta_t = g'(\pi)\frac{Q^i}{Q^t}^{t-L}$ for $L < t \leq L+M$. If $Q \geq \rho$, then $\zeta_t = g'(\pi)\frac{Q^i}{Q^t}^{t-1}$ for $t \leq L+M$.

We consider an infinite horizon model ($T = L = \infty$) for comparison with the sensitivity in the absence of manipulation, $\theta_t = (1-\rho)g'(\pi)$ from (21). $\zeta_t$ depends on whether $Q \leq \rho$.

\textsuperscript{24}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.
owing to the above trade-off arguments. If $Q < \rho$, manipulation 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 manipulation and need not change. If $Q > \rho$, the CEO does have incentives to manipulate under the original contract, and so the slope must increase to

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

The $(Q/\rho)^{t-1}$ term demonstrates that the slope 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 manipulate, and so the greater the required increase in sensitivity over time to deter manipulation. In a finite horizon model, $\theta_t$ is already increasing if manipulation is impossible; the feasibility of manipulation 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 manipulation on the contract. If $M = 1$, the contract changes from (22) to:

$$\ln c_1 = \frac{r_1}{5} + \kappa_1$$
$$\ln c_2 = \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2$$
$$\ln c_3 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3$$
$$\ln c_4 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{q_1r_4}{2} + \kappa_4$$
$$\ln c_5 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{q_1r_4}{2} + \kappa_5.$$

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 manipulation $q_1$; in the extreme, if $q_1 = 0$, manipulation 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 manipulate at $t = 1$ or $t = 2$ because there is no discounting, and so the negative effect of manipulation 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 manipulation forces the contract to change in $t \leq L$ even if there is no discounting.

### 6 Conclusion

This paper studies optimal CEO compensation in a dynamic setting in which the CEO consumes in each period, can privately save, and may temporarily manipulate returns. 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 manipulation is impossible, and rising if the horizon is finite or if manipulation is possible, even in the absence of career concerns. Deterring manipulation also requires the CEO to remain sensitive to firm returns after retirement. While the possibility of manipulation affects the sensitivity of pay, the option to privately save 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 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 manipulation.

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). We leave those extensions to future research.

\[\text{\underline{25}}\] 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 Dahya 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

As $L = T = \infty$ so that we have a constant $\theta_s = \theta$ and $k_s = k$. For notational simplicity we normalize (without loss of generality) $\bar{u} = 0$ and $\bar{a} = 0$. The expected cost of the contract is:

$$
C = E \left( \sum_{t=1}^{\infty} e^{-Rt} c_t \right) = \sum_{t=1}^{\infty} E \left[ \exp \left( -Rt + \ln c_0 + \sum_{s=1}^{t} \theta_s r_s + \sum_{s=1}^{t} k_s \right) \right]
$$

$$
= \sum_{t=1}^{\infty} \exp \left( (k - R + \ln E [e^{\theta \eta}]) t + \ln c_0 \right) = c_0 \frac{e^{k-R+\ln E[e^{\theta \eta}]} - 1}{1 - e^{k-R+\ln E[e^{\theta \eta}]}}
$$

The value of $c_0$ is pinned down by the participation constraint:

$$
0 = \bar{u} = E \left( \sum_{t=1}^{\infty} \rho^t \ln c_t \right) = \sum_{t=1}^{\infty} \rho^t \left( \ln c_0 + \sum_{s=1}^{t} \theta_s \bar{a} + \sum_{s=1}^{t} k_s \right) = \sum_{t=1}^{\infty} \rho^t \ln c_0 + kt
$$

$$
= \frac{\rho}{1 - \rho} \ln c_0 + \frac{\rho}{(1 - \rho)^2} k
$$

so that: $\ln c_0 = -\frac{1}{1-\rho^k}$. Hence

$$
C = e^{-\frac{1}{1-\rho}k} \frac{e^{k-R+\ln E[e^{\theta \eta}]} - 1}{1 - e^{k-R+\ln E[e^{\theta \eta}]}}
$$

For the contract without PS, we have $k = R + \ln \rho - \ln E [e^{\theta \eta}]$, so

$$
C^{\text{NPS}} = e^{-\frac{1}{1-\rho}(R+\ln \rho - \ln E [e^{\theta \eta}])} \frac{\rho}{1 - \rho}
$$

For the contract with PS, we have $k = R + \ln \rho + \ln E [e^{-\theta \eta}]$, so

$$
C^{\text{PS}} = e^{-\frac{1}{1-\rho}(R+\ln \rho + \ln E [e^{-\theta \eta}])} \frac{\rho e^{\ln E[e^{-\theta \eta}]+\ln E[e^{\theta \eta}]}}{1 - \rho e^{\ln E[e^{-\theta \eta}]+\ln E[e^{\theta \eta}]}}
$$
Thus,
\[
\Lambda = \frac{C^{PS}}{C^{NPS}} = \frac{(1 - \rho) e^{-\frac{\sigma}{1 - \delta}(\ln E[e^{-\theta \eta}] + \ln E[e^{\theta \eta}])}}{1 - \rho e^{\ln E[e^{-\theta \eta}] + \ln E[e^{\theta \eta}]}}
\]
\[
= \frac{1 - \rho}{1 - \rho e^{\Theta^2 \sigma^2} e^{-\frac{\sigma^2}{1 - \delta}}}
\]

In the limit of small time intervals, \( \ln E[e^{-\theta \eta}] + \ln E[e^{\theta \eta}] \sim \theta^2 \sigma^2 \), and \( 1 - \rho = \delta \) are small (proportional to the time interval \( \Delta t \)), and \( \theta \sim g'(\bar{a}) \delta \), so
\[
\Lambda \sim \frac{\delta e^{-\rho \theta^2 / (1 - \rho)}}{1 - (1 - \theta)(1 + \theta^2 \sigma^2)} \sim \frac{\delta e^{-\sigma^2 \theta^2 / (1 - \rho)}}{\delta - \theta^2 \sigma^2} = \frac{e^{-\sigma^2 \theta^2 / (1 - \rho)}}{1 - \frac{\theta^2 \sigma^2}{1 - \rho}}.
\]

### 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
\[
\ln c_t(r_1, ..., r_t) = \ln c_L(r_1, ..., r_L) + \kappa'_t,
\]
for some constants \( \kappa'_t \) 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
\[
\ln c_{t'}(r_1, ..., r_{t'}) = B(r_1, ..., r_t) + \theta_{t'} r_{t'} + \sum_{s=t+1}^{t-1} (\theta_s - \phi(s+1)) r_s + \kappa_{t'},
\]
for some function \( B \), constants \( \kappa_t \), and \( \theta_s \) as in the Theorem. The PS constraint yields
\[
c_t^{-\gamma} = e^{R \frac{J_{t+1}}{J_t} E_t c_t^{-\gamma}} = E_t \left[ e^{-\gamma B(r_1, ..., r_t) + R - \gamma \kappa_{t+1} + \ln J_{t+1} - \ln J_t} \right].
\]
We therefore have\(^{26}\)
\[
\ln c_t = B(r_1, ..., r_t) + \phi r_{t+1} + \kappa_t,
\]
for the appropriate constant \( \kappa_t \).

The EF constraint requires that
\[
0 \in \arg \max_{\varepsilon \geq 0} U(r_1, ..., r_{t-1}, \bar{\alpha} + \eta_t + \varepsilon).
\]

\(^{26}\)Equation (38) can also be derived from the IEE if \( \gamma = 1 \).
Since $g$ is differentiable, this yields (5) (see EG, Lemma 6), i.e.

$$J_t c_t^{1-\gamma} \phi_{t+1} + \frac{d}{d\varepsilon_-} B(r_1, \ldots, r_{t-1}, \bar{\alpha} + \eta_t + \varepsilon) \sum_{s=t}^T J_s E_t \left( c_s^{1-\gamma} \right) \geq J_t c_t^{1-\gamma} g'(\bar{\alpha}),$$

(40)

$$\frac{d}{d\varepsilon_-} B(r_1, \ldots, r_{t-1}, \bar{\alpha} + \eta_t + \varepsilon) \geq \sum_{s=t}^T J_s E_t \left( g'(\bar{\alpha}) - \phi_{t+1} \right) \to \sum_{n=t+1}^s E_t \left[ e^{(1-\gamma)\theta_n(\varepsilon_n+(1-\phi)\bar{\alpha})+(\kappa_n-\kappa_{n-1})} \right] = \theta_t - \phi_{t+1}.$$

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

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

We now show that (40) binds. First, (40) implies that for any $r' \geq r$ (see EG, Lemma 4)

$$B_t (r_1, \ldots, r_{t-1}, r') - B_t (r_1, \ldots, r_{t-1}, r) \geq (\theta_t - \phi_{t+1})(r' - r),$$

(41)

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

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

(42)

Condition (41) guarantees that the random variable $\ln c_t (r_1, \ldots, r_{t-1}, \tilde{r}_t)$ is weakly more dispersed than $\ln c_t^0 (r_1, \ldots, r_{t-1}, \tilde{r}_t).$ It also follows from the EF that both $\ln c_t (r_1, \ldots, r_{t-1}, \cdot)$ and $\ln c_t^0 (r_1, \ldots, r_{t-1}, \cdot)$ are weakly increasing. These facts, together with (42), imply that for the convex function $\psi$ and increasing function $\xi,$ where $\psi^{-1}(x) = e^{1-\gamma} x$, $\xi(x) = \frac{(1-\gamma)x}{1-\gamma}$ for $\gamma \neq 1$ and $\psi(x) = e^x$, $\xi(x) = x$ for $\gamma = 1,$ we have (see EG, Lemmas 1 and 2):

$$E_{t-1}[c_0^0(\tilde{r}_t, \ldots, r_t)] = E_{t-1} [\psi \circ \xi \circ \ln c_0^0(\tilde{r}_t, \ldots, r_t)] \leq E_{t-1} [\psi \circ \xi \circ \ln c_t(\tilde{r}_t, \ldots, r_t)] = E_{t-1} [c_t(\tilde{r}_t, \ldots, r_t)].$$

In the same way we show that $E_{t-1}[c_0^0(\tilde{r}_t, \ldots, r_s)] \leq E_{t-1}[c_0(\tilde{r}_t, \ldots, r_s)]$ for any $s \geq t.$ Consequently the contract $\{c_s^0 \}_{s \leq T}$ is cheaper than $\{c_s \}_{s \leq T},$ and so indeed (40) must bind.

Integrating out this equality we establish that for $t' \geq t,$

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

\footnote{Let $X$ and $Y$ denote two random variables with cumulative distribution functions $F$ and $G$ and corresponding right continuous inverses $F^{-1}$ and $G^{-1}.$ $X$ is said to be less dispersed than $Y$ if and only if $F^{-1}(\beta) - F^{-1}(\alpha) \leq G^{-1}(\beta) - G^{-1}(\alpha)$ whenever $0 < \alpha \leq \beta < 1.$}
where \( \theta_s \) are as required. Writing \( \kappa_0 = \ln c_0 \) and \( k_t = \kappa_t - \kappa_{t-1} \) establishes (25).

We now determine the values of the constants \( k_t \). First, we have \( c_0^{-\gamma} = e^{-\gamma \ln c_0} = e^{R_t J_t E \left[ c_t^{-\gamma} \right]} \) for \( t \leq T \) for all \( t \). This yields, for all \( t \):

\[
\gamma \sum_{s=1}^{t} k_s = R t + \ln J_t + \sum_{s=1}^{t} \ln E \left[ e^{-\gamma \theta_s (\varepsilon_s + (1 - \phi) \overline{\tau})} \right],
\]

yielding (27). When the PS constraint is not imposed, we use (7) to derive (10) analogously.

Equation (26) becomes simpler in the limit case \( L = T = \infty \). Then the problem is stationary, and \( \theta \) and \( k \) are constant. To characterize them, define \( f(\theta) = E \left[ e^{(1-\gamma)\theta (\varepsilon + \bar{a} (1 - \phi)) + k} \right] \) where \( \gamma k = R + \ln \rho + \ln E \left[ e^{-\gamma \theta (\varepsilon + \bar{a} (1 - \phi))} \right] \), so that

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

Then from (26), we have \( \theta = \frac{g'(\bar{a}) - \phi \theta}{\sum_{s=1}^{t} (\rho f(\theta))} + \phi \theta \), i.e.

\[
\theta = (g'(\bar{a}) - \phi \theta) (1 - \rho f(\theta)) + \phi \theta. \quad (43)
\]

In the limit of small time intervals, when \( \phi = 0 \), \( \theta \) satisfies:

\[
\theta = g'(\bar{a}) \left( - \ln \rho + \frac{\gamma - 1}{\gamma} (R + \ln \rho) + \frac{\gamma - 1}{2} (R - \ln \rho) \right)
\]

\[
= g'(\bar{a}) \left( \frac{(\gamma - 1) R - \ln \rho}{\gamma} + \frac{\gamma - 1}{2} (R - \ln \rho) \right)
\]

The value of \( \theta \) is the root that goes to a finite limit as \( \gamma \to 1 \):

\[
\theta = \frac{1 - \sqrt{1 - 2(\gamma - 1) \sigma^2 g'(\bar{a})^2 (\gamma - 1) (R - \ln \rho)}}{(\gamma - 1) \sigma^2 g'(\bar{a})} \quad (44)
\]

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:

\[
x_t = \begin{cases} 
- g(a_t) & \text{if } \gamma = 1 \\
 e^{-g(a_t)} \frac{1-\gamma}{\gamma} \beta & \text{if } \gamma \neq 1
\end{cases}
\]

\[
u(c_t, x_t) = \begin{cases} 
\ln c_t + x_t & \text{if } \gamma = 1 \\
\frac{c_t^{1-\gamma}}{1-\gamma} (3x_t)^{\gamma} & \text{if } \gamma \neq 1
\end{cases}
\]
where $\beta = \text{sign}(1 - \gamma)$, 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 $U \left( (c_t)_{t \leq T}, (x_t)_{t \leq L} \right) = \sum_{t=1}^{T} \beta^{t} u(c_t, x_t)$ be total discounted utility and consider the maximization problem:

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

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

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

for $f(x_s) = \bar{a}$ for $s > L$. Problems (45) and (31) are equivalent: $(x_t)_{t \leq L}$ and $(c_t)_{t \leq T}$ solve (45) if and only if $(f(x_t))_{t \leq L}$ and $(c_t)_{t \leq T}$ solve (31). 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 $f(x_t^*) = \bar{a}$), under no saving. For any path of noises $\eta = (\eta_t)_{t \leq T}$ we introduce the “upper linearization” utility function $\hat{U}_\eta$:

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

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:

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

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

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

when private savings are allowed, the PS constraint (6) implies that $e^{rt} \frac{\partial \hat{U}_\eta}{\partial c_t}$ is a martingale. Therefore, the agent is indifferent about when he consumes income $y_t$, and so we can evaluate $E\hat{U}_\eta$ 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). We also let utility be a function of \((x_t)_{t \leq T}\) since it 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 \(\tilde{U}_\eta((x_t)_{t \leq L}) = \tilde{U}_\eta((y_t)_{t \leq T}, (x_t)_{t \leq L})\) for \(\tilde{U}_\eta\) defined as in (47) and \(y_t\) as in (46), and consider the following maximization problem:

\[
\max_{x_t \text{ adapted}} E \left[ \tilde{U}_\eta((x_t)_{t \leq L}) \right]. \tag{48}
\]

If the target leisure level \((x^*_t)_{t \leq L}\) solves the maximization problem (48) then \((c^*_t)_{t \leq T}\) and \((x^*_t)_{t \leq L}\) solve the maximization problem (45).

**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 \(\tilde{U}_\eta\) is concave for every path of noises. (We will deal with the case \(\gamma \neq 1\) in step 4). We have:

\[
\tilde{U}_\eta((x_t)_{t \leq L}) = \sum_{t=1}^{T} \rho^t (\ln c^*_t(\eta) - 1) + \sum_{t=1}^{L} \rho^t x_t + \sum_{t=1}^{T} e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \overline{\sigma} - \phi(f(x_{s-1}) - \overline{\tau})) + t \ln \rho}. \tag{49}
\]

Joint concavity of (49) in \((x_t)_{t \leq L}\) is equivalent to the joint concavity of “PV of income” function

\[
I((x_t)_{t \leq L}) = \sum_{t=1}^{T} e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \overline{\sigma} - \phi(f(x_{s-1}) - \overline{\tau})) + t \ln \rho}. \tag{50}
\]

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 \(s:\)

\[
\sup [2C j'_s^2 + j''_s] \leq 0 \tag{51}
\]

for \(C = \sum_{n=0}^{T} e^{n \sup j'/2}\). 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 \(I((b_t)_{t \leq L})\) 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 \(t\), \(\exp(tj(b))\) is a convex function of \(b\), because its second derivative is \(\exp(tj(b))t(j'(b))^2 + j''(b))\), 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

\[
I([x_t]_{t \leq L}) = \sum_{t=1}^{T} e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \bar{\pi} - \phi(f(x_{s-1}) - \bar{\pi})) + t \ln \rho}
\]

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 \(\tilde{U}_\eta\) is:

\[
\tilde{U}_\eta([x_t]_{t \leq L}) = \sum_{t=1}^{L} \frac{\gamma}{1-\gamma} \rho^t e^\eta(\gamma) \left(\frac{x_t}{(\beta x_t^\gamma)^{1-\gamma}}\right) + \sum_{t=1}^{T} \rho^t (\beta x_t^\gamma) \epsilon_0 \epsilon_{\sum_{s=1}^{t} \theta_s (f(x_s) - \gamma \bar{\pi} - \phi(f(x_{s-1}) - \gamma \bar{\pi}) + (1-\gamma) \epsilon_s + (1-\gamma) k_s). \tag{52}
\]

Unlike when \(\gamma = 1\), the second term in (52), i.e. the “PV of income function”, now depends on noise \(\eta\). We therefore cannot prove pathwise concavity of linearized utility, and instead prove concavity of expected utility directly.

Expected utility is given by

\[
E\left[\tilde{U}_\eta([x_t]_{t \leq L})\right] = E \left[\sum_{t=1}^{L} A_t x_t + \sum_{t=1}^{T} M_t(\eta) e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \gamma \bar{\pi} - \phi(f(x_{s-1}) - \gamma \bar{\pi})) + (1-\gamma) \epsilon_s + (1-\gamma) k_s)} + t \ln \rho \right]
\]

\[
= E \left[\sum_{t=1}^{L} A_t x_t + M_T(\eta) \sum_{t=1}^{T} e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \gamma \bar{\pi} - \phi(f(x_{s-1}) - \gamma \bar{\pi})) + (1-\gamma) \epsilon_s + (1-\gamma) k_s)} + t \ln \rho \right],
\]

where \(M_t(\eta) = e^{\sum_{s=1}^{t} [(1-\gamma) \theta_s \epsilon_s - \ln E(e^{(1-\gamma) \theta_s \epsilon_s})]} + (1-\gamma) \ln c_0 - g(\bar{\pi})\) 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

\[
I'(([x_t]_{t \leq L}) = \sum_{t=1}^{T} e^{\sum_{s=1}^{t} \theta_s (f(x_s) - \gamma \bar{\pi} - \phi(f(x_{s-1}) - \gamma \bar{\pi})) + (1-\gamma) \epsilon_s + (1-\gamma) k_s)} + t \ln \rho,
\]

39
for $f(x_s) = \pi$ 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, $E\tilde{U}_\eta$ satisfies the first-order conditions at $(x_t^*)_{t \leq L}$. From step 4, $E\tilde{U}_\eta$ is also concave in $(x_t)_{t \leq L}$, and so the target leisure level $(x_t^*)_{t \leq L}$ solves the maximization problem (48). Therefore, from Lemma 1, $(c_t^*)_{t \leq T}$ and $(x_t^*)_{t \leq L}$ solve the maximization problem (45), establishing the result.

A.5 Proof of Theorem 4

We wish to show that, if baseline firm size $X$ is sufficiently large, the optimal contract implements maximum effort $(a_t \equiv \pi$ for all $t)$.

Fix any contract $(A,Y)$ that is incentive compatible and gives expected utility $u_\pi$ where $A = \{a_1, \ldots, a_L\}$ is the effort schedule, $a_t : [\eta, \pi]^T \rightarrow [0, \pi]$, and $Y = \{y_1, \ldots, y_T\}$ is the payoff schedule, $y_t : [\eta, \pi]^T \rightarrow \mathbb{R}$. The timing in each period is as follows: the agent reports noise $\eta_t$, then is supposed to exert effort $a_t(\eta_1, \ldots, \eta_t)$. If the return is $y_t + a_t(\eta_1, \ldots, \eta_t)$ he receives payoff $y_t(\eta_1, \ldots, \eta_t)$, else he receives a payoff that is sufficiently low to deter such "off-equilibrium" deviations. We require this richer framework, since in general the noises might not be identifiably from observed returns (when $y_t + a_t(\eta_1, \ldots, \eta_t) = y_t' + a_t(\eta_1, \ldots, \eta_{t-1}, \eta_t')$ for $\eta_t \neq \eta_t'$). Note that the required low payoff may be negative. A limited liability constraint would be simple to address, e.g. by imposing a lower bound on $\eta$.

To establish the result it is sufficient to show that we can find a different contract $(A^*, Y^*)$ that implements maximum effort $(a_t \equiv \pi$ for all $t)$, and is not significantly costlier than $(A,Y)$, in the sense that

$$
E \left[ \sum_{t=1}^{T} e^{-\gamma t} (y_t^*(\eta_t) - y_t(\eta_t)) \right] \leq h(E[\pi - a_1(\eta_1)], \ldots, E[\pi - a_L(\eta_L)]),
$$

(53)

for some linear function $h$, $h : \mathbb{R}^L \rightarrow \mathbb{R}$, with $h(0, \ldots, 0) = 0$. This is sufficient, because if initial firm size $X$ is sufficiently large, then for every sequence of noises and actions, firm value $X e^{\sum_{s=1}^{t-1}(y_s + a_s(\eta_s)) + \pi}$ is greater than $D$, where $D$ is the highest slope coefficient of $h$. This in turn implies

$$
X e^{\sum_{s=1}^{t-1}(y_s + a_s(\eta_s)) + \pi} \times E \left[ e^\pi - e^{a_t(\eta_t)} \right] \geq D \times E \left[ \pi - a_t(\eta_t) \right],
$$

(54)

and so the benefits of implementing maximum effort outweigh the costs, i.e. the RHS of (53) exceeds the LHS of (53). To keep the proof concise we assume $p e^\pi = 1$, $T = L$ and the noises $\eta_t$ are independent across time. The general case is proven along analogously.

We introduce the following notation. For any contract $(A,Y)$ and history $\eta_t$, let $u_t(\eta_t) = \frac{y_t(\eta_t) e^{-g(a_t(\eta_t))}}{1-\gamma}$ (or $u_t(\eta_t) = \ln y_t(\eta_t) - g(a_t(\eta_t))$ for $\gamma = 1$) denote the CEO’s stage game utility for truthful reporting in period $t$ after history $\eta_t$ when he consumes his income, let $U_t(\eta_t) = E_t \left[ \sum_{s=t}^{L} \rho^{s-t} u_s(\eta_s) \right]$ denote his continuation utility, and $m u_t(\eta_t) = y_t^\gamma(\eta_t) e^{-(1-\gamma)g(a_t(\eta_t))}$ denote his marginal utility of consumption. We divide the proof into the following six steps.
Step 1. Local necessary conditions. First, we generalize the local effort constraint (5) to contracts that need not implement maximum effort.

**Lemma 5** Fix an incentive compatible contract \((A, Y_0)\), with each \(a_t(\eta_{t-1}, \cdot)\) continuous almost everywhere and bounded on every compact subinterval, and a history \(\eta_{t-1}\). The CEO’s continuation utility \(U_t(\eta_{t-1}, \eta_t)\) must satisfy the following:

\[
U_t(\eta_{t-1}, \eta_t) = U_t(\eta_{t-1}, \eta_t) + \int_{n} [y_t(\eta_{t-1}, x)e^{-g(a_t(\eta_{t-1}, x))}]^{1-\gamma} g'(a_t(\eta_{t-1}, x))dx, \tag{55}
\]

with \(y_t(\eta_t) > 0\).

Step 2. Bound on the cost of incentives per period. For any history \(\eta_{t-1}\) and contract \((A, Y_0)\), consider “repairing” the contract at time \(t\) as follows. Following any history \(\eta_{t-1}, \eta_t\), multiply all the payoffs by the appropriate constant \(\zeta(\eta_{t-1}, \eta_t)\) such that the continuation utilities \(U^\#_t(\eta_{t-1}, \eta_t)\) for the resulting contract satisfy (55) with \(a_t(\eta_{t-1}, \eta_t) = \pi\) for all \(\eta_t\). In other words, the local EF constraint for maximum effort at time \(t\) after history \(\eta_{t-1}\) is satisfied. The following Lemma bounds the expectation of how much we have to scale up the payoffs by the expectation of how much the target effort falls short of the maximum effort.

**Lemma 6** Fix an incentive compatible contract \((A, Y_0)\) and a history \(\eta_{t-1}\), and consider the contract \((A^\#, Y^\#)\) such that:

\[
a_t^\#(\eta_{t-1}, \eta_t) = \pi \text{ for all } \eta_t, \text{ else } a_s^\# \equiv a_s, \]

\[
y_s^\#(\eta_s) = y_s(\eta_s) \times \zeta(\eta_{t-1}, \eta_t) \text{ if } \eta_s|_{t} = \eta_{t-1}, \eta_t, \text{ and else } y_s^\#(\eta_s) \equiv y_s(\eta_s),
\]

where \(\zeta(\eta_{t-1}, \eta_t) \geq 1\) is the unique number such that \(U^\#_t(\eta_{t-1}, \eta_t) = U_t(\eta_{t-1}, \eta_t)\) and

\[
U^\#_t(\eta_{t-1}, \eta_t) = U_t(\eta_{t-1}, \eta_t) + \int_{n} [\zeta(\eta_{t-1}, x) y_t(\eta_{t-1}, x)e^{-g(\pi)}]^{1-\gamma} g'(\pi)dx. \tag{56}
\]

Then:

\[
E_{t-1} [\zeta(\eta_{t-1}, \eta_t)] \leq \varphi(E_{t-1} [\pi - a_t(\eta_t)]), \tag{57}
\]

where \(\varphi(x) = e^{g'(\pi)\sup \frac{\mu}{\text{f}} \gamma < 1} \left(1 + e^{g(\pi)} - g(\pi)(1 - \gamma)x\right)\) for \(\gamma \neq 1\),

\[
\varphi(x) = e^{g'(\pi)\sup \frac{\mu}{\text{f}}} (1 + e^{g(\pi)} - g(\pi)x) \text{ for } \gamma = 1, \text{ and } f \text{ is the pdf of noise } \eta.
\]

Step 3. Constructing the contract that satisfies the local EF constraint in every period. We want to use the procedure from step 2 to construct a new contract \((A^x, Y^x)\) that requires maximum effort, satisfies the local EF in every period, and has a cost difference over \((A, Y_0)\) that is bounded by how much \((A, Y_0)\) falls short of maximum effort. For this we need the following Lemma.
Lemma 7 For a contract \((A, Y)\) and any \(\zeta > 0\) consider the contract \((A, \zeta Y)\) in which all the payoffs are multiplied by \(\zeta\),

i) if \((A, Y)\) satisfies the local EF constraint then so does \((A, \zeta Y)\);

ii) if \((A, Y)\) satisfies the local PS constraint then so does \((A, \zeta Y)\).

Given an incentive compatible contract \((A, Y)\), we construct the contract \((A^x, Y^x)\) as follows. The contract always prescribes maximum effort. Regarding the payoffs, for any period \(t\) after a history \(\eta_{t-1}\) we first multiply all payoffs after history \((\eta_{t-1}, \eta)\) with fixed constants \(\zeta(\eta_{t-1}, \eta) > 1\) as in Lemma 6 so that the resulting utilities \(U^\#_t(\eta_t)\) satisfy (56). Then we multiply all payoffs following history \(\eta_{t-1}\) by the appropriate constant \(\zeta^{\mu t}(\eta_{t-1}) < 1\) so that for the resulting contract \((A^x, Y^x)\) we obtain the original promised utility, i.e. \(U_{t-1}(\eta_{t-1}) = U^x_{t-1}(\eta_{t-1})\). By construction and the above Lemmas, the contract \((A^x, Y^x)\) satisfies the local EF constraint. In particular, due to Lemma 7, repairing the contract after history \(\eta_{t-1}\) will not upset the local EF constraint after history \((\eta_{t-1}; \eta_t)\).

The original contract \((A, Y)\) satisfies the local PS constraint, i.e. the current marginal utility of consumption always equals the next-period expected marginal utility. Providing incentives for maximum effort in contract \((A^x, Y^x)\) upsets this condition. In the following two steps, given \((A^x, Y^x)\), we construct the contract \((A^\#, Y^\#)\) that also satisfies the local PS constraint and is not much costlier. In particular, we show that the extent to which the marginal utilities of consumption in \((A^\#, Y^\#)\) depart from the marginal utilities in \((A^x, Y^x)\) is bounded by the extent to which effort falls short of maximum effort in contract \((A, Y)\).

Step 4. Bound on the decrease of expected MU of consumption per period. We split this step into two Lemmas. The first bounds the expected decrease in marginal utility of consumption from providing incentives for maximum effort in the current period, as in step 2. The second bounds the decrease in expected marginal utility by the expected decrease of the marginal utility.

Lemma 8 Fix any history \(\eta_{t-1}\) and look at the original contract \((A, Y)\) and the contract \((A^\#, Y^\#)\) from step 1. Then:

\[
E_{t-1} \left[ \begin{array}{c}
\frac{mu^\#_t(\eta_{t-1}, \eta_t)}{mu^h_t(\eta_{t-1}, \eta_t)}
\end{array} \right] \geq e^{-\gamma g(\pi)} \sup_{\eta_t} \frac{g''}{g} E_{t-1} \left[ \pi - a_t(\eta_t) \right] \left( 1 - 1_{\gamma < 1} e^{-(1+\gamma)(1-\gamma)} \right) g'(\pi) \left( 1 - \gamma \right) \left( 1 + \gamma \right) E_{t-1} \left[ \pi - a_t(\eta_t) \right].
\]

Lemma 9 Fix any history \(\eta_{t-1}\) and look at any two contracts \((A^i, Y^i)\) \((A^h, Y^h)\) with positive payoffs that satisfy (55) and for every \(\eta_t\), \(mu^i_t(\eta_{t-1}, \eta_t) \leq mu^h_t(\eta_{t-1}, \eta_t)\). Then, for some \(D_2 > 0\):

\[
\frac{E_{t-1} \left[ mu^i_t(\eta_{t-1}, \eta_t) \right]}{E_{t-1} \left[ mu^h_t(\eta_{t-1}, \eta_t) \right]} \geq 1 - D_2 \left( 1 - E_{t-1} \left[ \frac{mu^i_t(\eta_{t-1}, \eta_t)}{mu^h_t(\eta_{t-1}, \eta_t)} \right] \right).
\]
Step 5. **Constructing the contract that satisfies the local PS constraint in every period.** Providing incentives for maximum effort in \((A^*, Y^*)\) at (say) time \(L\) affects the marginal utility of consumption in period \(L\) and upsets the PS constraint in period \(L-1\). However, restoring the PS constraint in period \(L-1\) will affect the marginal utility of consumption in period \(L-1\) and so upset the PS constraint in period \(L-2\), and so on. In the following Lemma we bound this overall effect using Lemma 8 and iteratively Lemma 9.

**Lemma 10** There is a contract \((A^*, Y^*)\) that implements maximum effort and satisfies the local EF and PS constraints, and for every history \(\eta_t\):

\[
\frac{\mu_2^*(\eta_t)}{\mu_2^I(\eta_t)} \geq \prod_{s=t+1}^{L} \phi^{s-t}(E_t[\psi(E_{s-1}[\bar{\alpha} - a_s(\eta_s)])]),
\]

where \(\phi(x) = 1 - D_2 \left(1 - x\right), \psi(x) = e^{-\gamma g'(\bar{\alpha})} \sup_{g''} \left(1 - 1_{\gamma < 1} e^{-(1+\gamma)(1-\gamma)g(\bar{\alpha}) - g'(\bar{\alpha})}(1 - \gamma)(1 + \gamma)x\right)\).

Step 6. **Bounding the cost difference (53).** By construction, contract \((A^*, Y^*)\) from Lemma 10 implements maximum effort, causes the local EF constraint to bind, satisfies the local PS constraint and leaves the CEO with the expected discounted utility \(u\). Therefore it is identical to the contract from Theorem 2, and so also satisfies the global constraints (Theorem 3). It therefore remains to prove (53).

One can verify that for some \(D_3 > 0\) for every history \(\eta_t\) we have \(y_t^*(\eta_t) < D_3\). Moreover, for any \(a, b, c \in \mathbb{R}\),

\[
a - b \leq a \left(\max\{\frac{a-c}{c}, 0\} + \max\{\frac{c - b}{b}, 0\}\right) = a \left(\max\{\frac{a}{c}, 1\} - 1 + \max\{\frac{c}{b}, 1\} - 1\right).
\]

Consequently,

\[
E \left[\sum_{t=1}^{L} e^{-rt} (y_t^*(\eta_t) - y_t(\eta_t))\right] \leq D_3 \times E \left[\sum_{t=1}^{L} e^{-rt} \left(\max\{\frac{y_t^*(\eta_t)}{y_t(\eta_t)}, 1\} - 1 + \max\{\frac{y_t^*(\eta_t)}{y_t(\eta_t)}, 1\} - 1\right)\right] \leq D_3 \times E \left[\sum_{t=1}^{L} e^{-rt} \left(\prod_{s=t+1}^{L} \phi^{s-t}(E_t[\psi(E_{s-1}[\bar{\alpha} - a_s(\eta_s)])])\right)^{-\frac{1}{L}} - 1 + \varphi \left(E_{t-1}[\bar{\alpha} - a_t(\eta_t)]\right) - 1\right],
\]

where \(\varphi\) is as in Lemma 6, while \(\phi\) and \(\psi\) are as in Lemma 10. All functions \(\varphi, \phi, \psi, \prod_{s=t+1}^{L} x_s\) and \(x^{-\frac{1}{L}}\) are continuously differentiable and take value 1 for argument(s) equal to 1, whereas \(\bar{\alpha} - a_0(\eta_t)\) is bounded. Therefore there is a linear function \(h : \mathbb{R}^L \rightarrow \mathbb{R}\) with \(h(0, ..., 0) = 0\) such that (53) is satisfied.

The above proof is for the case where private saving is possible as this is the more complex case. If \(\gamma = 1\) and private saving is impossible, step 4 is not needed and Lemma 10 in step 5 and step 6 become significantly simpler.
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 Theorem, and $\theta_t$ deterministic lowest nonnegative values such that the EF and NM constraints are satisfied, i.e.:

$$EF : g'(\bar{a}) \leq \theta_t \left(1 + \rho + \ldots \rho^{T-t}\right) \text{ for } t \leq L,$$  \hspace{1cm} (59)

$$NM : \theta_t (\rho^t + \ldots + \rho^T) q_i \leq \rho^{t+i} \theta_t (\rho^{t+i} + \ldots + \rho^T), \text{ for } 0 \leq t \leq L, 0 \leq i \leq M. \hspace{1cm} (60)$$

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

$$g'(\bar{a}) \leq \zeta_t \text{ for } t \leq L$$

$$\zeta_t q_i \leq \rho^i \zeta_{t+i} \text{ for } 0 \leq t \leq L, 0 \leq i \leq M$$

This yields the values described in the Proposition.
References


