Dynamic Incentive Accounts[*]

Alex Edmans   Xavier Gabaix   Tomasz Sadzik   Yuliy Sannikov

Wharton   NYU and NBER   NYU   Princeton

March 17, 2009

Preliminary and incomplete - for conference presentation only

Abstract

Optimal contracts in a dynamic model must address a number of issues absent from static frameworks. The CEO can manipulate earnings in the short run; he may undo contracts through private saving; and shocks to firm value can weaken the incentive effect of securities over time. We analyze the optimal compensation scheme in such a setting. The efficient contract takes a surprisingly simple form, and can be implemented by a "Dynamic Incentive Account". The CEO's expected pay is escrowed into an account, a fraction of which is invested in the firm's stock and the remainder in cash. The account features state-dependent rebalancing and time-dependent vesting. It is constantly rebalanced so that the equity fraction remains above a certain threshold; this threshold sensitivity is typically increasing over time even in the absence of career concerns. The account vests gradually both during the CEO's employment and after he quits, to deter short-termist actions before retirement.

KEYWORDS: Contract theory, executive compensation, incentives, principal-agent problem, manipulation, private saving, vesting.
JEL CLASSIFICATION: D2, D3, G34, J3

1 Introduction

Many classical models of CEO compensation consider only a single period, or multiple unlinked periods. However, the optimal contract in a static analysis may be suboptimal in a dynamic world where the CEO's current actions, such as his effort or savings/consumption choice, impact future periods. For example, short-term contracts can encourage the CEO to manipulate earnings or scrap investment projects to boost the current stock price at the expense of long-run value. By privately saving, the CEO can separate his consumption stream from the path of income provided by his contract, and thus "undo" the intended intertemporal incentives. The incentive effect of firm securities may change over time, so that they no longer have the initially desired effect: if firm value declines, options may fall out-of-the-money and bear little sensitivity to the stock price. In addition to the three above challenges, a dynamic setting also provides opportunities absent from a static framework – in particular, the firm has the option to reward current effort with future rather than contemporaneous compensation.

This paper analyzes optimal executive compensation in a dynamic model that allows for all of the above complexities, which are likely important features in real life. In our framework, the CEO consumes in multiple periods, thus allowing current effort to be compensated in the future, but may choose to save rather than simply consuming the income provided by the contract. He can temporarily boost current earnings through manipulation or earnings smoothing; the long-term costs may not appear until after the CEO has retired. Furthermore, firm value is subject to shocks in each period which will affect the value, and incentive effect, of any securities the CEO is given as part of his incentive contract.

In an infinite horizon model where the CEO has no option to manipulate earnings or privately save, the optimal contract is time-independent: the sensitivity of pay to the firm's return is the same in each period. The optimal contract is also scale-independent. In our model, the relevant measure of incentives is the percentage change in CEO pay for a percentage change in firm value; translated into real variables, this is the fraction of CEO pay that comprises of stock. If the CEO's outside option doubles, his total pay doubles but the relative weighting on cash and stock remains the same. This result extends to a dynamic setting Edmans, Gabaix and Landier (2009), who advocated this incentive measure in a

one-period model with a risk-neutral CEO. The optimal contract also involves consumption smoothing. Since the agent is risk-averse, it is efficient to spread the reward for effort across all future periods rather than concentrating it in the current period (the "deferred reward principle"). This result is consistent with Boschen and Smith (1995), who find that changes in firm value has a much greater effect on future rather than contemporaneous pay.

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

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

By contrast, the possibility of private savings does not change the sensitivity of the contract to firm value, since it does not affect the CEO's action. Instead, giving the CEO the option to privately save impacts the time trend of the level of pay. This optimal contract now involves a greater rise in compensation over time compared to a setting in which saving is impossible. Rising pay is necessary to deter the CEO from wishing to save today to increase future consumption.

Despite the complications that result from a dynamic setting, the optimal contract can be implemented in a surprisingly simple manner. When initially appointed, the CEO is given a "Dynamic Incentive Account": a portfolio of which a given fraction is invested in the firm's stock and the remainder in cash. As time evolves, and firm value changes, this portfolio is continuously rebalanced, so that the fraction in the firm's stock remains sufficient to induce effort at minimum risk to the CEO. This fraction represents the contract's sensitivity, and so

is constant in an infinite horizon model where manipulation is impossible, and increasing over time otherwise. For example, a fall in the share price decreases the equity in the incentive account; this is addressed by using cash in the account to purchase stock. By contrast, if the share price rises, some of the equity can be sold to reduce the risk borne by the CEO.

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

In sum, the Dynamic Incentive Account has two key features, which each achieve separate objectives. State-dependent rebalancing ensures that the CEO always has sufficient incentives to exert effort, while minimizing the risk that he bears. Time-dependent vesting dissuades the CEO from manipulating earnings, while allowing him to finance consumption. In this paper, vesting and rebalancing are separate events. In most real-life compensation schemes, vesting and rebalancing are one and the same event – the CEO can only sell his securities for cash when they vest. Thus, while placing long vesting periods on stock and options deters myopia, it does not achieve rebalancing and thus maintain their incentive effect as the share price changes.

Similarly, existing theories of vesting horizons analyze vesting and rebalancing as being the same event. Peng and Roell (2009) derive the optimal vesting period as a trade-off: distant vesting deters manipulation but increases the risk borne by the CEO as it delays rebalancing of stock for cash. Brisley (2006) and Bhattacharyya and Cohn (2008) show that allowing the CEO to rebalance his securities for cash can increase his willingness to undertake risky projects by reducing his firm-specific risk. Since rebalancing can only be achieved through vesting, Bhattacharyya and Cohn show that the optimal vesting period is short. While they consider stock, Brisley analyzes options where rebalancing is only necessary upon strong performance, since only in-the-money options subject the CEO to risk.[1] Therefore, as

---

[1] Another difference between their models is that in Brisley (2006), the securities vest before investment decisions are made; in Bhattacharrya and Cohn (2008) vesting occurs afterwards.

4

in our model, state-dependent rebalancing is optimal; in Brisley, rebalancing must coincide with vesting and so this entails state-dependent vesting. Indeed, recent empirical studies (e.g. Bettis, Bizjak, Coles and Kalpathy (2008)) document that performance-based (i.e. state-dependent) vesting is becoming increasingly popular. However, state-dependent vesting may allow the CEO to manipulate the stock price upwards and cash out his shares. Thus, state-dependent vesting has critically different effects to the combination of state-dependent rebalancing and time-dependent vesting. Our framework incorporates manipulation and so requires these two features to achieve the two separate goals of effort inducement and manipulation deterrence.

In addition to the above papers on vesting horizons, our paper is also related to the literature on optimal contracts in the presence of manipulation. Lacker and Weinberg (1989) identify a class of settings in which no manipulation is optimal and linear contracts obtain. Goldman and Slezak (2006) model the trade-off between effort inducement (which increases the optimal equity stake) and manipulation deterrence (which reduces it). More generally, the theory is related to dynamic models of the principal-agent problem, such as DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), He (2008a), Sannikov (2008) and Garrett and Pavan (2009). Our modeling setup bears some similarities to the multi-period framework of Edmans and Gabaix (2008) ("EG"), who also generate time-independent contracts. Technically, we draw the "detail-neutral" features of our contracts from that paper: for instance, the functional form of the contract is independent of the noise distribution and agent's utility function. However, EG do not consider manipulation and restrict the CEO to consuming in the final period only. The most closely related paper is He (2008b) who considers a dynamic setting in which the agent can privately save and also engage in a myopic action (similar to manipulation in this paper). He shows that the optimal wage pattern is non-decreasing over time, that sufficiently good past performance leads to permanent pay raises, and that severance pay is efficient. Our model has quite different specifications (multiplicative utility, continuous action choice and a cost function that need not be linear) which leads to a scale-independent, closed-form contract. Our analysis focuses on the state-dependent rebalancing and time-dependent vesting of the dynamic incentive account.

In addition to its results, our paper contributes a number of methodological innovations.

5

To our knowledge, it is the first to derive conditions on the model primitives which guarantee the validity of the first-order approach to solve a dynamic agency problem with private savings. An agency problem is a maximization problem subject to the agent's incentive constraints. The first-order approach replaces the incentive constraints against complex multi-period deviations with weaker local constraints (i.e. first-order conditions), with the hope that the solution to the relaxed problem satisfies all incentive constraints.[2] This method is often valid without private savings (hence the one-shot deviation principle), but it has proved problematic when the agent can save. The difficulties arise since the agent can engage in joint deviations to save and reduce effort, because savings provide insurance against future shocks to income and thus reduce the agent's incentives to exert effort in the future. Our method to guarantee the validity of the first-order approach centers around viewing the agent's total lifetime income as a function of his total disutility of effort. If this function is concave, the first-order approach is valid, since the agent's utility is concave in income.

The bulk of the analysis derives the (exactly) optimal contract to implement a given effort level, as in Grossman and Hart (1983). We then endogenize the optimal path of effort levels, and this extension contains our second methodological innovation. Following the argument of Fong and Sannikov (2009) that predictions of optimal contracting theories are important only insofar as they have sizable, rather than negligible, impact on profitability, we aim to derive a simple contract that is close to the optimal contract in terms of efficiency, rather than *the* complicated optimal contract. Our contract is *approximately* optimal under the assumption that firm value is significantly larger than the CEO's wage, which is indeed true in practice. Under this assumption, the difference in profitability between our contract and the optimal contract converges to 0 as firm's earnings become larger. The methodological innovation here is in proving that the contract is approximately optimal without deriving the optimal contract. To do so, we construct an upper bound on profit that any contract can attain, justify it using martingale methods, and show that our simple contract comes

---

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

close to the upper bound. (See also He (2008b) who uses a related technique in a different setting.)

This paper is organized as follows. Section 2 presents the discrete time version of the model and derives the optimal contract when the CEO has logarithmic utility, as this version of the model is most tractable. We show that the contract involves consumption smoothing and typically rising incentives over time, and that it can be implemented in practice using the Dynamic Incentive Account. Section 3 shows that the key economics of the contract continue to hold under general CRRA utility functions, autocorrelated noise and in continuous time. Section 4 concludes.

2 The Core Model

2.1 Assumptions

We consider a multiperiod model featuring a firm (also referred to as the principal) which offers a contract to a CEO (also referred to as the agent). The CEO's utility function is given by:

$$
U = \begin{cases} \sum_{t=1}^{T} e^{-\rho t} \frac{\left(c_t e^{-g(a_t)}\right)^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \sum_{t=1}^{T} e^{-\rho t} \left(\ln c_t - g\left(a_t\right)\right) & \text{if } \gamma = 1, \end{cases} \tag{1}
$$

where $e^{-\rho t}$ represents a discount factor, $\gamma > 0$ is the CEO's relative risk aversion, c_t is consumption and a_t is effort (also referred to as "action"), defined over some interval $[\underline{a}_t, \overline{a}_t]$. The action is broadly defined to encompass any decision that improves firm value but is personally costly to the manager. The main interpretation is effort, but it can also refer to rent extraction, in which case a low a_t reflects cash flow diversion or private benefit consumption.

The utility function in (1) exhibits multiplicative preferences, i.e. the effect of effort on the CEO's utility depends on his level of consumption. Such preferences are common in macroeconomic models and consider private benefits as a normal good, consistent with the treatment of most goods and services in consumer theory. Edmans, Gabaix and Landier (2009) show that multiplicative preferences lead to scale-independent contracts. The CEO's

reservation utility is $\underline{u}$.

The CEO works until time L and then retires. He lives until time $T \geq L$. If $T > L$, the CEO takes no action from time $L + 1$ to T (i.e. $a_t = 0$) but continues to consume. In each period t, a signal of the CEO's effort is released. The core interpretation of the signal is the firm's stock return and so we will use the terms "signal" and "return" interchangeably; other interpretations will be discussed later. In the absence of manipulation, the signal is given by

$$R_t = a_t + \eta_t, \tag{2}$$

where $\eta_1, ..., \eta_T$ are independent noises and $\eta_2, ..., \eta_T$ have log-concave densities.[3] Section 3.1 extends the model to autocorrelated noises.

As in EG, we assume that in each period t, the CEO first observes the noise η_t and then takes action a_t, before observing the noise in the next period. This timing assumption is also featured in models in which the CEO observes the "state of nature" before choosing his effort level, as well as cash flow diversion models where the CEO sees total output before deciding how much to steal (e.g. DeMarzo and Fishman (2007)). EG show that this assumption leads to tractable contracts in discrete time, as well as consistent results with the continuous time case, where noise and actions are simultaneous.

Manipulation of Returns. We allow for the CEO to manipulate the firm's return. In practice, such manipulation can take many forms. In the most literal interpretation, the manager can change accounting policies to accelerate the realization of revenues or delay the impact of costs (either by concealing information, or capitalizing rather than expensing costs).[4] Alternatively, he can engage in short-termist behavior by scrapping investment projects (as modeled by Stein (1988)) or taking on risky projects (such as sub-prime lending) for which the potential downside may not manifest for several years. In both cases, the increase in current earnings is at the expense of future profits. Note that manipulation may

---

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

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

be downwards as well as upwards: if the contract involves rising bonuses over time (as in Gibbons and Murphy (1992)), the CEO may be tempted to sacrifice current earnings to boost future profits. In practice, this can be achieved by investing in negative-NPV projects, or "big bath" accounting (taking large write-downs in the current period).

We model such manipulation as follows. At each time t, the CEO can choose to engage in manipulation m_t, and simultaneously selects the "release lag" $i_t \in \{1, ..., M\}$, which is the interval before the manipulation becomes apparent. For example, forgoing an investment project that pays off in the very long-run will only worsen earnings far into the future, and so the release lag i_t is high. In the presence of manipulation, the firm's returns are

$$r_t = R_t + m_t - \lambda(m_t), \qquad r_{t+i_t} = R_{t+i_t} - m_t, \qquad r_s = R_s \text{ for } s \neq t, t+i_t, \qquad (3)$$

where λ represents the cost of manipulation. We assume $\lambda(0) = \lambda'(0) = 0$, and $\lambda(m) > 0$ for $m \neq 0$. Therefore, manipulation lowers returns at period $t + i_t$ by m_t, and only boosts returns in period t by $m_t - \lambda(m_t)$. Manipulation is inefficient owing to the deadweight cost $\lambda(m_t)$. In reality, such losses arise because resources are required to change accounting policies, positive-NPV projects are being scrapped (for $m_t > 0$), or negative-NPV projects are being pursued (for $m_t < 0$). The principal observes r_t in each period, but not a_t, m_t or η_t.

Private Saving. We allow for the CEO to privately save or borrow, so that he can separate his consumption stream from the path of income provided by the contract. If the CEO saves d_t at time t, he invests it in a bank account yielding an instantaneous risk-free interest rate r_f, and thus can consume an additional $d_t e^{r_f s}$ at time $t + s$. We allow for d_t to be negative: the CEO may consume out of past savings, or borrow (i.e. he may have negative savings). The optimal contract must be robust to private saving.

The Contract. The contract is a set of functions $c_t(r_1, ..., r_t)$: $\mathbb{R}^t \rightarrow \mathbb{R}$, for $t > 0$. We do not require the contract to be linear or to be implementable with standard securities (e.g. stock and options). EG show that independent and log-concave noises, combined with a non-increasing absolute risk aversion (NIARA) utility function, are sufficient to rule out stochastic contracts and contracts that require the agent to send messages to the principal. Since the utility function is CRRA, we have NIARA. Therefore, we can restrict the analysis

9

to deterministic, message-free contracts.

The Firm's Objective. As in Grossman and Hart (1983), Dittmann and Maug (2007) and EG, we fix the path of effort levels that the principal wants to induce. In each period t, the principal wishes to implement (at least) a_t^*, where $a_t^* > \underline{a}_t$ and a_t^* is allowed to be time-varying. EG show in a one-period model that, if the firm is large enough, a_t^* will equal the maximum effort level $\bar{a}_t$. This is because the benefits of effort are a function of firm size, and the costs of effort (both direct disutility and the inefficient risk-sharing caused by using an incentive contract to induce effort) are proportional to the CEO's wage. If the firm is sufficiently large compared to the CEO's wage, the benefits of effort swamp the costs, and maximum effort is efficient. A maximum effort level will exist because there is a limit either to the number of productive activities that a CEO can undertake (e.g., finite NPV-positive projects) or to the number of hours in a day the CEO can work while remaining productive. (Under the interpretation of a as rent extraction, the maximum effort level reflects zero stealing.) We revisit this "maximum effort principle" formally in section 3.3. For the remainder of the paper, we will assume that $a_t^* = \bar{a}_t$.

We also assume that the principal wishes to deter all manipulation ($m_t = 0 \; \forall \; t$). Manipulation is costly because it is inefficient ($\lambda(m_t) \geq 0$); the benefit of allowing manipulation is that it may permit a less sensitive and thus cheaper contract. Intuitively, if the firm is sufficiently large compared to the CEO, the costs of manipulation exceed the benefits and zero manipulation is optimal. Appendix A.5 contains potential microfoundations for the optimality of zero manipulation.

The firm wishes to find the cheapest contract that ensures the CEO's participation and satisfies up to three constraints. The first is the standard "incentive compatibility" (IC) constraint, which ensures that the CEO exerts (at least) the desired effort level in each period, i.e. $a_t \geq a_t^* \; \forall \; t$. Second, the "no manipulation" (NM) constraint ensures that the CEO does not engage in costly manipulation, i.e. $m_t = 0 \; \forall \; t$. Third, the "private savings" (PS) constraint ensures that the CEO does not wish to undo the contract via private savings, i.e. $d_t = 0 \; \forall \; t$.

To highlight the effect of allowing private savings and manipulation on the optimal contract, we will consider versions of the model in which manipulation and/or private saving

are ruled out.

2.2 Local Constraints

Our solution strategy is as follows. First, we find the best contract among all contracts that satisfy the local constraints. Second, we verify that this contract satisfies all constraints, i.e. the agent will not wish to undertake global deviations. We start with the first stage: this section derives the local constraints, which we use to derive the contract in Section 2.3 . Later, we will prove that this contract also deters global deviations.

Contract $c_t\left(r_1, ..., r_t\right)$ yields the following utility:

$$
U\left(r_1, ..., r_T\right) = \begin{cases} \sum_{t=1}^{T} e^{-\rho t} \frac{\left(c_t(r_1,...,r_t)e^{-g(a_t)}\right)^{1-\gamma}}{1-\gamma} & \text{if } \gamma \neq 1 \\ \sum_{t=1}^{T} e^{-\rho t}\left(\ln c_t\left(r_1, ..., r_t\right) - g\left(a_t\right)\right) & \text{if } \gamma = 1. \end{cases} \tag{4}
$$

$c_t\left(r_1, ..., r_t\right)$ is a stochastic variable whose value depends on the past history of signal realizations, but for conciseness we will suppress this dependence and write c_t. If the CEO takes the recommended actions $\left(a_t^*\right)_{t\geq 0}$, his utility is:

$$
U = \begin{cases} \sum_{t=1}^{T} B_t c_t^{1-\gamma}/\left(1-\gamma\right) & \text{if } \gamma \neq 1 \\ \sum_{t=1}^{T} B_t\left(\ln c_t - g\left(a_t^*\right)\right) & \text{if } \gamma = 1, \end{cases} \tag{5}
$$

where

$$
B_t = e^{-\rho t - (1-\gamma)g(a_t^*)}. \tag{6}
$$

We first address the IC constraint and consider a local deviation ε from the target action a_t^*. If the CEO exerts effort $a_t = a_t^* + \varepsilon$, the firm's return increases from $r_t = \eta_t + a_t^*$ to $r_t = \eta_t + a_t^* + \varepsilon$. The CEO's utility rises by

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

This deviation should be non-positive for $\varepsilon < 0$. We therefore require

$$
E_t\left[\frac{\partial U}{\partial r_t}\frac{\partial r_t}{\partial a_t} + \frac{\partial U}{\partial a_t}\right] \geq 0.
$$

11

Since $\frac{\partial r_t}{\partial a_t} = 1$ and $\frac{\partial U}{\partial a_t} = -g'(a_t^*) B_t c_t^{1-\gamma}$, the IC constraint is:

$$IC : E_t \left[\frac{\partial U}{\partial r_t} \right] \geq g'(a_t^*) B_t c_t^{1-\gamma} \text{ for } 0 \leq t \leq L. \tag{7}$$

We next address the NM constraint. If the CEO undertakes a small manipulation m_t, the return in period t becomes $r_t = R_t + m_t - \lambda(m_t)$. For some $i \leq M$, the return in period $t + i$ becomes $r_{t+i} = R_t - m_t$. His utility rises by

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

To prevent manipulation, this increase in utility must be zero. Since $\lambda(0) = \lambda'(0) = 0$, the $\lambda(m_t)$ term drops out for small manipulations. Hence, the NM constraint is:

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

Finally, we consider the PS constraint. Since private saving does not affect the manager's action, we can ignore the disutility of effort $(g(a_t))$ and focus solely on the positive utility generated by income. Let $V(c_1, ..., c_T)$ denote this utility. If the CEO saves d_t in period t and invests it until $t + s$, his utility increases by

$$-E_t \left[\frac{\partial V}{\partial c_t} \right] d_t + E_t \left[\frac{\partial V}{\partial c_{t+s}} \right] d_t e^{r_f s}.$$

To deter private saving (or dis-saving), this change should be zero, i.e.

$$E_t \left[e^{r_f t} \frac{\partial V}{\partial c_t} \right] = E_t \left[e^{r_f (t+s)} \frac{\partial V}{\partial c_{t+s}} \right],$$

the Euler equation. Therefore, the PS constraint is that $e^{r_f t} \frac{\partial V}{\partial c_t}$ is a martingale. Intuitively, if it were not a martingale, the agent would privately save to reallocate consumption to the time periods in which marginal utility is higher. For the utility function (5), this becomes:

$$PS: \frac{B_t c_t^{-\gamma}}{e^{-r_f t}} \text{ is a martingale.} \tag{9}$$

This condition can be contrasted with the "Inverse Euler Equation" (IEE), which charac-

terizes a vast set of agency problems without the PS constraint (Rogerson (1985), Golosov, Kocherlakota and Tsyvinski (2003) and Farhi and Werning (2009)):

$$IEE: \frac{e^{-r_f t}}{B_t c_t^{-\gamma}} \text{ is a martingale.} \tag{10}$$

The intuition for the Inverse Euler Equation is as follows. The inverse of the agent's marginal utility is the marginal cost of delivering utility to the agent at a given moment of time. Equation (10) is the first-order condition for giving the agent a given level of utility in the cheapest possible way. If (10) did not hold, the principal could benefit by shifting the agent's consumption to periods with a lower value of $\frac{e^{-r_f t}}{B_t c_t^{-\gamma}}$.

Without private saving, the principal chooses the time path of consumption so to minimize the marginal cost of providing utility. With private saving, the agent chooses the time path of consumption to maximize its marginal utility, i.e. maximize the reciprocal of the marginal cost. This explains why the IEE is the inverse of the PS constraint.

2.3 Optimal Contract, Log Utility

We now derive the optimal contract. We first present the contract under log utility, as the expressions are most transparent and the key principles are the same as the general CRRA case. Section 3.1 considers the general CRRA case, as well as extends the model to autocorrelated noise.

Theorem 1 (*Optimal contract, log utility*). *The optimal contract that satisfies the local constraints pays the CEO c_t in period t, where c_t satisfies:*

$$\ln c_t = \sum_{s=1}^{t} \theta_s r_s + \kappa_t, \tag{11}$$

and θ_s and κ_t are deterministic functions. Without the NM constraint:

$$\theta_t = \begin{cases} \frac{1-e^{-\rho}}{1-e^{-\rho(T-t+1)}} g'(a_t^*) & \text{for } t \leq L \\ 0 & \text{for } t > L. \end{cases} \tag{12}$$

13

With the NM constraint,

$$\theta_t = \begin{cases} \frac{(1-e^{-\rho})e^{\rho t}}{1-e^{-\rho(T-t+1)}}\Theta & \text{for } t \le L+M \\ 0 & \text{for } t > L+M, \end{cases} \qquad (13)$$

where $\Theta = \sup_{s \le L} (e^{-\rho s} g'(a_s^*))$. Let $\zeta = 1$ denote the case when private savings are allowed (and so the PS constraint is imposed) and $\zeta = -1$ if they are ruled out (and so the PS constraint is not imposed). The value of κ_t is given by:

$$\kappa_t = (r_f - \rho)t + \zeta \ln E\left[e^{-\zeta \sum_{s \le t} \theta_s r_s}\right] + \underline{\kappa} \text{ for } t \le T, \qquad (14)$$

where $\underline{\kappa}$ is chosen to ensure that the agent is at his reservation utility:

$$\underline{u} = \sum_t e^{-\rho t} \left(\underline{\kappa} + (r_f - \rho)t + \zeta \ln E\left[e^{-\zeta \sum_{s \le t} \theta_s r_s}\right] + E\left[\sum_{s \le t} \theta_s r_s\right] - g(a_t^*) \right).$$

Proof (Heuristic). The Appendix presents a formal proof. Here, we provide a heuristic proof that conveys the "essence" of the economic argument. We consider the case of $L = T = 2$ and use the following reasoning from EG. (7) yields: $e^{-2\rho} d(\ln c_2)/dr_2 \ge e^{-2\rho} g'(a_2^*)$. In the Appendix we show that the cheapest contract involves this local IC condition binding, i.e. $d(\ln c_2)/dr_2 = g'(a_2^*) \equiv \theta_2$. Integrating yields the contract:

$$\ln c_2 = \theta_2 r_2 + B(r_1), \qquad (15)$$

where $B(r_1)$ is a function of r_1. It is the "constant" viewed from time 2.

For brevity, we consider only the case without the NM constraint. The case with the NM constraint is proven similarly but rather more complex in discrete time; the arguments can be seen more clearly in the continuous-time heuristic proof in Section 3.2. If the PS constraint is not imposed, we use the IEE (10). Applying this for $t = 1$ gives:

$$c_1 = e^{\rho - r_f} E_1[c_2] = E\left[e^{\theta_2 r_2}\right] e^{B(r_1) + \rho - r_f}. \qquad (16)$$

If the PS constraint is imposed, we apply PS (9) for $t = 1$ to give:

$$c_1 = e^{r_f - \rho} E_1[c_2] = E\left[e^{\theta_2 r_2}\right] e^{B(r_1) + r_f - \rho}.$$

14

In both cases, we obtain

$$\ln c_1 = B\left(r_1\right) + k, \tag{17}$$

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

$$U = e^{-\rho}\ln c_1 + e^{-2\rho}\ln c_2 + k' = \left(e^{-\rho} + e^{-2\rho}\right)B\left(r_1\right) + k''. \tag{18}$$

We next apply (7) to (18) to yield: $\left(e^{-\rho} + e^{-2\rho}\right)B'\left(r_1\right) \geq e^{-\rho}g'\left(a_1^*\right)$. Again, the cheapest contract involves this condition binding, i.e. $\left(e^{-\rho} + e^{-2\rho}\right)B'\left(r_1\right) = e^{-\rho}g'\left(a_1^*\right)$. Integrating yields:

$$B\left(r_1\right) = \theta_1 r_1 + k''', \tag{19}$$

where $\theta_1 = e^{-\rho}g'\left(a_1^*\right)/\left(e^{-\rho} + e^{-2\rho}\right)$. Combining (19) with (15) yields:

$$\ln c_2 = \theta_1 r_1 + \theta_2 r_2 + \kappa_2,$$

for some constant κ_2. Combining (19) with (17) yields:

$$\ln c_1 = \theta_1 r_1 + \kappa_1,$$

for another constant κ_1.

We finally determine the values of the constants κ_t. Since this part of the proof is equally clear for general T as for $T = 2$, we show it for the general case. When the PS constraint is not imposed, we use the IEE (10). There exists a value $e^{\underline{\kappa}}$ such that $e^{\underline{\kappa}} = E\left[\frac{e^{-r_f t}c_t}{e^{-\rho t}}\right]$ for all t. This yields, for all t,

$$e^{\underline{\kappa}} = \frac{e^{-r_f t + \kappa_t}}{e^{-\rho t}}E\left[e^{\sum_{s=1}^{t}\theta_s r_s}\right] = \frac{e^{-r_f t + \kappa_t}}{e^{-\rho t}}\prod_{s=1}^{t}E\left[e^{\theta_s r_s}\right],$$

i.e. (14) with $\zeta = -1$.

When the PS constraint is imposed, we use (9). There exists a value $e^{-\underline{\kappa}}$ such that $e^{-\underline{\kappa}} = E\left[\frac{e^{r_f t}e^{-\rho t}}{c_t}\right]$ for all t. This yields $e^{-\underline{\kappa}} = e^{-\rho t}e^{r_f t - \kappa_t}E\left[e^{-\sum_{s=1}^{t}\theta_s r_s}\right] = e^{-\rho t}e^{r_f t - \kappa_t}\prod_{s=1}^{t}E\left[e^{-\theta_s r_s}\right]$, i.e. (14) with $\zeta = 1$. ∎

15

The remaining step is to show that the agent will not wish to undertake global deviations, e.g. jointly reducing effort and saving. Since this proof is equally clear for general γ as for log utility, we delay the proof until Section 3.1.2 which demonstrates the result for general CRRA utility functions.

We now discuss the economics behind the optimal contract. (11) shows that time-t consumption should be linked not only to the signal in period t, but also the signals in all previous periods. Therefore, exerting effort in a particular period boosts income not only in that period, but also in all future periods. We call this the *"deferred reward principle"*: since the CEO is risk-averse, it is optimal to spread the reward for effort across all future periods rather than concentrate it in the period in which effort is exerted. This prediction is consistent with Boschen and Smith (1995), who find that changes in firm value has a much greater effect on future rather than contemporaneous pay.

We now consider how the contract sensitivity changes over time. (11) shows that, in an infinite horizon model ($T = \infty$) with a constant target action ($a_t^* = a^*$), the sensitivity of the contract is constant and given by:

$$\theta_t = \theta = \left(1 - e^{-\rho}\right) g'(a^*). \tag{20}$$

The time-independent sensitivity is intuitive: the contract must be sufficiently sharp to compensate for the disutility of effort, and the latter is constant when the target action does not change over time. However, in a finite model, the contract's sensitivity is increasing over time, even if the target action is constant. The intuition for this *"increasing incentives principle"* is similar to the above deferred reward principle: there are fewer remaining periods over which to smooth out the reward for effort, and so the CEO must earn a greater reward in each period. As in Gibbons and Murphy (1992), our model generates the prediction that CEOs closer to retirement must have sharper contracts. While Gibbons and Murphy obtain this result by invoking career concerns, our explanation is that consumption smoothing possibilities decline towards retirement.

Next, we study the impact of manipulation on the optimal contract. The possibility of manipulation has three main effects. First, it ensures that the CEO remains sensitive to the stock price after his retirement in period L: he remains sensitive until period $L + M$. This

16

is to deter him from manipulating the signal just before departure. Second, it causes the contract sensitivity to be higher in each period (compared to a case in which manipulation is impossible), because the contract must now satisfy the NM constraint as well as the IC constraint. Third, it affects how the contract sensitivity trends over time. If this sensitivity were time-independent, the CEO would have an incentive to manipulate the time-t return upwards, thus increasing his time-t consumption. Even though the return at time $t + i_t$ will be lower, the effect on the CEO's consumption is smaller in present value terms owing to discounting. Therefore, an increasing sensitivity is necessary to deter manipulation. For example, in an infinite horizon model ($T = \infty$) with a constant target action ($a_t^* = a^*$), the possibility of manipulation changes the contract from the constant (20) to

$$\theta_t = \left(1 - e^{-\rho}\right) e^{\rho t} g'\left(a^*\right)$$

The $e^{\rho t}$ term demonstrates the increasing slope. The more impatient the CEO, the greater the incentives to manipulate, and so the greater the required increase in sensitivity over time to deter manipulation. In a finite horizon model, the slope is already increasing if manipulation is ruled out; the possibility of manipulation causes it to rise even faster. I

With a constant target action, the effect of the NM constraint on the speed with which the contract's sensitivity rises over time depends only on the CEO's impatience ρ. With a non-constant target action, it depends on $\Theta = \sup_{t \leq L} \left(e^{-\rho t} g'\left(a_t^*\right)\right)$ the maximum discounted sensitivity during the CEO's working life. Let $s \leq L$ denote the period in which $e^{-\rho t} g'\left(a_t^*\right)$ is highest. The CEO has an incentive to increase r_s at the expense of the signal in any t within M periods of s. Therefore, the sensitivity for all t within M periods of s must increase, to remove these incentives. However, this in turn has a knock-on effect: since the sensitivity for $t = s - M$ has now risen, the CEO now has an incentive to increase r_{s-M} at the expense of r_{s-2M}, and so on. Therefore, the sensitivity at s forces upward the sensitivity in all periods $t \leq L + M$, even those more than M periods away from s, because of the knock-on effects. This explains why the contract in all periods $t \leq L + M$ depends on Θ in equation (13).

Finally, the possibility of private savings affects the constant κ_t but not the sensitivity of the contract θ_t. Since private saving does not affect the agent's action, the optimal sensitivity of CEO pay to the action is unchanged. Instead, the possibility of private saving affects the

time trend of the contract. The constant κ_t is given by (14). When private savings are allowed ($\zeta = 1$), the second term in (14) declines more slowly over time than if private savings are ruled out ($\zeta = -1$). Thus, the need to deter private savings leads to CEO pay having a greater upward trend over time than in the absence of this constraint. This result is consistent with He (2008b), who finds that the optimal contract under private savings involves a wage pattern that is non-decreasing over time.

2.4   Optimal Contract, Log Utility, Numerical Example

This section uses a simple numerical example to show most clearly the deferred reward and increasing incentives principles, as well as the effect of manipulation on the optimal contract. We first set $T = 5$, $L = 3$, $\rho = 0$ and $g'(a_t^*) = 1$ for all t, and assume that manipulation is impossible. Applying L'Hopital's rule to (12), the optimal contract is given by:

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

This example shows both principles at work. First, there is consumption smoothing: an increase in r_1 augments log consumption (i.e. the CEO's utility) in all future periods by the same amount. Second, the sensitivity increases over time, from 1/5 to 1/4 to 1/3. Since the CEO takes no action from $t = 4$ onwards, his consumption does not depend on r_4 or r_5. However, his consumption at $t = 4$ and $t = 5$ continues to depend on r_1, r_2 and r_3 as his earlier efforts affect his wealth, from which he consumes until death.

If the CEO can manipulate earnings, the contract changes to:

$$\ln c_1 = \frac{r_1}{5} + \kappa_1$$

$$\ln c_2 = \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2$$

$$\ln c_3 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3$$

$$\ln c_4 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{r_4}{2} + \kappa_4$$

$$\ln c_5 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{r_4}{2} + \kappa_5.$$

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

If either of the above conditions are violated, then the contract must change in all periods when manipulation is possible. First, we allow for discounting by changing ρ to 0.1, and keeping all other parameters constant. The optimal contract is now

$$\ln c_1 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \kappa_1$$

$$\ln c_2 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{1 - e^{-0.1}}{1 - e^{-0.4}} r_2 + \kappa_2$$

$$\ln c_3 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{1 - e^{-0.1}}{1 - e^{-0.4}} r_2 + \frac{1 - e^{-0.1}}{1 - e^{-0.3}} r_3 + \kappa_3$$

$$\ln c_4 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{1 - e^{-0.1}}{1 - e^{-0.4}} r_2 + \frac{1 - e^{-0.1}}{1 - e^{-0.3}} r_3 + \kappa_4$$

$$\ln c_5 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{1 - e^{-0.1}}{1 - e^{-0.4}} r_2 + \frac{1 - e^{-0.1}}{1 - e^{-0.3}} r_3 + \kappa_5$$

under no manipulation. If manipulation is possible with $M = 1$, the contract changes to

$$\ln c_1 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \kappa_1$$

$$\ln c_2 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{\left(1 - e^{-0.1}\right) e^{0.1}}{1 - e^{-0.4}} r_2 + \kappa_2$$

$$\ln c_3 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{\left(1 - e^{-0.1}\right) e^{0.1}}{1 - e^{-0.4}} r_2 + \frac{\left(1 - e^{-0.1}\right) e^{0.2}}{1 - e^{-0.3}} r_3 + \kappa_3$$

$$\ln c_4 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{\left(1 - e^{-0.1}\right) e^{0.1}}{1 - e^{-0.4}} r_2 + \frac{\left(1 - e^{-0.1}\right) e^{0.2}}{1 - e^{-0.3}} r_3 + \frac{\left(1 - e^{-0.1}\right) e^{0.3}}{1 - e^{-0.2}} r_4 + \kappa_4$$

$$\ln c_5 = \frac{1 - e^{-0.1}}{1 - e^{-0.5}} r_1 + \frac{\left(1 - e^{-0.1}\right) e^{0.1}}{1 - e^{-0.4}} r_2 + \frac{\left(1 - e^{-0.1}\right) e^{0.2}}{1 - e^{-0.3}} r_3 + \frac{\left(1 - e^{-0.1}\right) e^{0.3}}{1 - e^{-0.2}} r_4 + \kappa_5.$$

When manipulation is possible, not only do c_4 and c_5 now depend on r_4, but also only does the possibility of manipulation mean that r_4 affects the CEO's consumption, but it also the contract's sensitivity increases more rapidly between $t = 1$ and $t = 3$. Since the CEO is impatient, the old contract gives him an incentive to sacrifice future returns for current earnings. Therefore, a more rapidly increasing slope is needed so that future returns must have a greater effect on his consumption to remove these incentives.

Second, we revert to $\rho = 0$ and instead vary the marginal cost of effort by setting $g'(a_1^*) = g'(a_2^*) = 1$ and $g'(a_3^*) = 3$. If manipulation is impossible, the optimal contract is

$$\ln c_1 = \frac{r_1}{5} + \kappa_1$$

$$\ln c_2 = \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2$$

$$\ln c_3 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3}r_3 + \kappa_3$$

$$\ln c_4 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3}r_3 + \kappa_4$$

$$\ln c_5 = \frac{r_1}{5} + \frac{r_2}{4} + \frac{2}{3}r_3 + \kappa_5.$$

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

$$\ln c_1 = \frac{2}{5}r_1 + \kappa_1$$

$$\ln c_2 = \frac{2}{5}r_1 + \frac{r_2}{2} + \kappa_2$$

$$\ln c_3 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + \kappa_3$$

$$\ln c_4 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + r_4 + \kappa_4$$

$$\ln c_5 = \frac{2}{5}r_1 + \frac{r_2}{2} + \frac{2}{3}r_3 + r_4 + \kappa_5.$$

2.5 Implementation of the Optimal Contract: the Dynamic Incentive Account

Taking first differences of (11) yields:

$$\ln c_t - \ln c_{t-1} = \theta_t r_t + \kappa_t - \kappa_{t-1}. \tag{21}$$

The contract thus prescribes the percentage change in CEO pay ($\ln c_t - \ln c_{t-1}$) as a function of the firm's return r_t, i.e. the percentage change in firm value. The relevant measure of incentives is therefore the elasticity of CEO pay to firm value; this elasticity must be at least θ_t to ensure incentive compatibility. Empiricists have used a number of statistics to measure incentives – for example, Jensen and Murphy (1990) calculate "dollar-dollar" incentives (the dollar change in CEO pay for a dollar change in firm value) and Hall and Liebman (1998) estimate "dollar-percent" incentives (the dollar change in CEO pay for a percentage firm return.) By contrast, Murphy (1999) advocates the elasticity measure ("percent-percent" incentives) on empirical grounds: it is invariant to firm size, and firm returns have much greater explanatory power for percentage than dollar changes in pay. However, he notes that "elasticities have no corresponding agency-theoretic interpretation." The above analysis provides a theoretical justification for using elasticities to measure incentives. Edmans, Gabaix and Landier (2009) showed that percent-percent incentives are the optimal measure if effort has a multiplicative effect on both CEO utility and firm value.[5] Their result was derived in a one-period model with a risk-neutral CEO; we extend it to a dynamic model with a risk-averse CEO who can manipulate the stock price and privately save. Our setting contains the above two features: the utility function in (1) exhibits multiplicative preferences, and effort has an additive effect on the firm's percentage return (equation (2)) and thus a multiplicative effect on firm value. In terms of real variables, percent-percent incentives equal the fraction of total pay that is comprised of stock. The required fraction (θ_t) is independent of total pay: if the CEO's outside option doubles, total pay doubles. Therefore, the value of equity must double to ensure that the fraction of total pay invested in equity remains the same – the fraction is scale-independent.

To ensure that percent-percent incentives equal θ_t in each period t, the contract can be implemented in the following manner. The present value of the CEO's expected pay is escrowed into a "Dynamic Incentive Account" ("DIA") at the start of period $t = 1$.[6] A

---

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

[6] We present one possible implementation of the optimal contract; other implementations are possible.

proportion θ_1 of the Incentive Account is invested in the firm's stock and the remainder in cash. At the start of each subsequent period t, this portfolio is rebalanced so that the proportion invested in the firm's stock is θ_t. This dynamic rebalancing addresses a common problem of option compensation: if firm value declines, the option's delta falls and so its incentive effect is reduced. Unrebalanced stock compensation suffers from the same problem, even though the delta of a share is 1 regardless of firm value. The relevant measure of incentives is not the delta of the CEO's portfolio (which represents the dollar change in CEO wealth for a dollar change in firm value) but the proportion of CEO wealth which is in firm shares (which represents percent-percent incentives). When the stock price falls, the value of the CEO's shares declines but his cash is unaffected. Therefore, stock constitutes a smaller proportion of the CEO's wealth, which reduces his incentives. The DIA addresses this problem by exchanging cash for stock, to maintain the fraction of stock in the account at θ_t. Importantly, the additional stock is accompanied by a reduction in cash – it is not given for free. This addresses a major concern with repricing options after stock price declines to restore incentives – the CEO is rewarded for failure. By contrast, if the stock price rises, the value of the stock increases and so becomes a higher fraction of the account. Therefore, some of his shares can be sold for cash (thus reducing the CEO's risk) without incentives falling below θ_t. Indeed, Fahlenbrach and Stulz (2008) find that decreases in CEO ownership typically occur after good performance.

The DIA thus features dynamic rebalancing to ensure that the IC constraint is satisfied in each period. This rebalancing is state-dependent: if the stock price rises (falls), stock is sold (bought) for cash. The second key feature of the DIA is time-dependent vesting: the CEO can only withdraw a fraction α_t of the account in each period for consumption (we will later derive α_t in a specific case. This gradual vesting ensures that the NM constraint is satisfied in each period: it prevents the CEO from manipulating earnings and then cashing out his entire equity before the manipulation is discovered. Moreover, vesting is gradual not only during the CEO's tenure but also after retirement. The CEO is not paid the entire DIA in period L. Instead, it only fully vests in period $L + M$, to ensure that all manipulation has been

---

For example, rather than placing the entire present value of the CEO's future pay in the account at the start, only his $t = 1$ reservation wage could be invested initially. In each subsequent period, the reservation wage of that period is added to the account.

reversed. Therefore, the CEO's income remains sensitive to firm returns after retirement, to prevent him from acting myopically (via manipulation or taking large risks) just before his departure. Commentators have argued that the latter problem was particularly important in the recent financial crisis. For example, Angelo Mozilo, the former CEO of Countrywide Financial, made over $100m from stock sales prior to his firm's collapse; a November 20, 2008 *Wall Street Journal* article entitled "Before the Bust, These CEOs Took Money Off the Table" provides further examples. More broadly, Johnson, Ryan and Tian (2009) find a positive correlation between corporate fraud and unrestricted (i.e. immediately vesting) stock compensation.

In sum, the DIA has two key features. Time-dependent vesting ensures that the CEO does not manipulate earnings, and allows him to smooth consumption. State-dependent rebalancing guarantees that the CEO has sufficient incentives to exert effort, while minimizing the risk that he bears. Some existing compensation schemes satisfy the first feature, but not the second. For example, restricted stock and options vest along a given time schedule, irrespective of firm performance. Long-vesting securities are effective in satisfying the NM constraint but not the IC constraint – if firm value falls, they represent a smaller percentage of the CEO's wealth and so have a weaker incentive effect. Hence, the DIA is critically different from the restricted securities observed empirically.

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

effort incentives.

We demonstrate the workings of the DIA in an infinite horizon model ($T = \infty$) with a constant target action ($a_t^* = a^*$). The optimal contract sensitivity is constant and given by (20). The CEO's consumption is:

$$c_t = c_0 e^{\theta R_t + Gt}, \text{ where } G \equiv r_f - \rho + \zeta \ln E\left[e^{-\zeta\theta(a^*+\eta)}\right]. \tag{22}$$

Let $A_0 = E_0\left[\sum_{t=1}^{T} e^{-r_f t} c_t\right]$ be the initial value of the DIA, i.e. the present value of future earnings. A fraction θ is invested in the firm's stock and the remainder in cash, so that the account evolves according to: $dA_t/A_t = (r_f - \alpha)\,dt + \theta\sigma dZ_t$. The CEO withdraws a fraction α of the account in each period, so that his consumption is $c_t = \alpha A_t$. This is intuitive, since an agent with log utility wishes to consume a constant fraction of his wealth in each period, and this fraction is independent of the rate of return on his wealth. From (22), we obtain $\alpha = \rho - (1 + \zeta)\sigma^2\theta^2/2$.

If the PS constraint is not imposed, we have $\alpha = \rho$ and the inverse marginal utility, c_t, is a martingale so that the agent does not wish to reallocate consumption across time periods to increase his marginal utility. Elementary calculations lead to $A_0 = e^{\underline{\kappa}}/\left(e^{\delta} - 1\right)$. If the PS constraint is imposed, we have $\alpha < \rho$. The agent would like to invest zero wealth in the stock as it carries a zero risk premium, but he is forced to invest θ and bear unrewarded risk. Therefore, the agent will wish to save to insure himself against this risk. To remove these incentives, we must have $\alpha < \rho$ so that the account grows faster than it vests, thus providing automatic saving for the agent. We also have $A_0 = e^{\underline{\kappa}}/\left(e^{\delta - \ln E\left[e^{-\theta(a^*+\eta)}\right] - \ln E\left[e^{+\theta(a^*+\eta)}\right]} - 1\right)$.

3 Extensions

This section analyzes extensions to the core model. Section 3.1 considers autocorrelated signals and general CRRA utility functions, and Section 3.2 studies continuous time. The economics of the optimal contract in Section 2 are robust to both extensions.

3.1 General CRRA Utility and Autocorrelated Signals

The core model assumes that the signal r_t was the firm's stock return. This is an attractive interpretation for a number of reasons: it allows the optimal contract to be implemented using the firm's securities, and it allows us to assume that the noises η_t are uncorrelated. However, in private firms, there is no stock return, and so alternative signals of effort must be used such as profits. Unlike stock returns, shocks to profits may be serially correlated. This subsection extends the model to such a case.

We now assume that the noises $\eta_1, ..., \eta_T$ follow an $AR(1)$ process with autoregressive parameter ϕ, i.e. $\eta_t = \phi\eta_{t-1} + \varepsilon_t$, $\phi \in [0,1]$, where ε_t are independent with an interval support $(\underline{\varepsilon_t}, \overline{\varepsilon_t})$ and the bounds may or may not be finite. To simplify the proofs, we make the following technical assumption:

$$g'(a^*_{t-1}) \geq \phi g'(a^*_t), \text{ for } t \leq L. \tag{23}$$

We also now extend the model to allow for a general CRRA utility function.

3.1.1 Optimal Contract

Theorem 2 *The optimal contract pays the CEO c_t in period t, where c_t satisfies:*

$$\ln c_t = \sum_{s=1}^{t} \theta_s r_s + \chi_t r_t + \kappa_t, \tag{24}$$

for deterministic constants θ_s, χ_t and κ_t. χ_t is defined inductively as:

$$\chi_t = \phi(\theta_{t+1} + \chi_{t+1}) \text{ for all } t. \tag{25}$$

Without the NM constraint,

$$\theta_t = \chi_t = 0 \text{ for } t > L, \tag{26}$$

$$\theta_t = \frac{B_t\left(g'(a^*_t) - \chi_t\right)}{\sum_{s=t}^{T} B_s e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a^*_t)} E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^{s}(\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^{s}\theta_m a^*_m + \chi_s a^*_s\right]}\right)} \text{ for } t \leq L.$$

With the NM constraint,

$\theta_t = \chi_t = 0$ *for* $t > L + M$,

$$\theta_t = \frac{e^{(1-\gamma)(\kappa_{L+M}-\kappa_t-\chi_t a_t^*)} E_t \left[e^{(1-\gamma)\sum_{m=t+1}^{L+M}(\theta_m+\chi_m)\varepsilon_m + \sum_{m=t+1}^{L}\theta_m a_m^*}\right] \times D - B_t \chi_t}{\sum_{s=t}^{T} B_s e^{(1-\gamma)(\kappa_s-\kappa_t-\chi_t a_t^*)} E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^{s}(\theta_m+\chi_m)\varepsilon_m + \sum_{m=t+1}^{s}\theta_m a_m^* + \chi_s a_s^*\right]}\right)} \quad \text{for } t \le L + M.$$

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

$$\gamma\kappa_t = \underline{\kappa} + r_f t + \ln B_t + \zeta \sum_{s=1}^{t} \ln E\left[e^{-\zeta\gamma(\theta_s+\chi_s)(a_s^*-\phi a_{s-1}^*+\varepsilon_s)}\right] \quad \text{for } t \le T, \tag{27}$$

where $\underline{\kappa}$ *is chosen to ensure that the agent is at his reservation utility:*

$$\sum_{t=1}^{T} \frac{\exp\left(-\rho t + (1-\gamma)\left(\sum_{s=1}^{t}\theta_s r_s + \chi_t r_t + \kappa_t - g(a_t)\right)\right)}{1-\gamma} = \underline{u},$$

and D *is the lowest constant such that:*

$$e^{(1-\gamma)(\kappa_{L+M}-\kappa_t-\chi_t a_t^*)} E_t \left[e^{(1-\gamma)\sum_{m=t+1}^{L+M}(\theta_m+\chi_m)\varepsilon_m + \sum_{m=t+1}^{L}\theta_m a_m^*}\right] \times D \ge B_t g'(a_t^*), \text{ for all } t \le L.$$

Proof See Appendix. ∎

Taking first differences of (24) and using (25) yields:

$$\ln c_t - \ln c_{t-1} = (\theta_t + \chi_t)(r_t - \phi r_{t-1}) + \kappa_t - \kappa_{t-1}. \tag{28}$$

We can therefore see the effect of allowing for general CRRA utility functions and autocorrelated noise. With independent noise ($\phi = 0$), $\chi_t = 0$ and so contracts (24) and (28) reduce to (11) and (21). Therefore, moving from log to CRRA utility but retaining independent noise (i.e. continuing to interpret r_t as the stock return) has little effect on the functional form of the optimal contract. As before, the contract links the percentage change in CEO pay to the absolute signal in period t – i.e. the percentage change in firm value. Hence, "percent-percent" incentives remain optimal. The consumption smoothing and increasing incentive

27

principles continue to hold, and the contract can be implemented using a dynamic incentive account. The qualitative effect of the NM and PS constraints on the optimal contract is the same as in Section 2. The key difference is that the parameters θ and κ are somewhat more complex.

In the presence of autocorrelated signals, the χ_t's are no longer zero. From (28), the optimal contract now links the percentage change in CEO pay in period t to innovations in the signal $(r_t - \phi r_{t-1})$ between t and $t-1$, rather than the absolute signal in period t. This is intuitive: since good luck (i.e. a positive shock) in the last period carries over to the current period, the contract should control for the last period's signal to avoid paying the CEO for luck. In particular, if returns follow a random walk, and if also $a_t^* = a^*$ and $T = L$ for simplicity, then the contract has a constant sensitivity:

$$\ln c_t = g'(a^*)r_t + \kappa_t$$

and (28) becomes

$$\ln c_t - \ln c_{t-1} = g'(a^*)(r_t - r_{t-1}) + \kappa_t - \kappa_{t-1}.$$

The percentage change in pay is linked to the absolute innovation in the signal. High profits in period t do not improve compensation if profits were equally high in the previous period.

3.1.2 Global Constraints

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

To consider global deviations, we must distinguish between the CEO's income and his consumption. The contract in Theorem 2 pays the agent an income y_t, given by

$$y_t = \exp\left(\sum_{s=1}^{t} \theta_s(\eta_s + a_s) + \chi_t(\eta_t + a_t) + \kappa_t\right), \tag{29}$$

where the constants θ_s, χ_t and κ_t are as in (25), (26) and (27). Without the PS constraint, the agent's consumption c_t simply equals his income y_t. When the agent can privately save, these two variables are distinct.

The following Theorem shows that if the cost function g is sufficiently convex, the CEO has no profitable global deviation. Consider the following assumption:

$$CONV : 4 \sup g'^2 \left(\min \left\{ \frac{e^{(\sup \chi_t (\bar{a}_t - a_t^*) + \sup \chi_t (a_t^* - \underline{a}_t))/2}}{1 - \min\{\delta, 1\}}, T \right\} + \max\{\frac{1-\gamma}{\gamma}, 0\} \right) - \inf g'' < 0 \tag{30}$$

where

$$\delta = \exp \left\{ \sup_s \left\{ -\rho + \theta_s(\bar{a}_s - \gamma a_s^*) + (1-\gamma) \left[(\kappa_s - g(a_s^*)) - (\kappa_{s-1} - g(a_{s-1}^*)) + \ln E(e^{\theta_s e_s}) \right] \right\} / 2 \right\}. \tag{31}$$

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

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

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

with $\sum_{t=1}^T e^{-r_f t} (y_t - c_t) \geq 0$ and y satisfying (29). Assume $\gamma = 1$ or $\alpha = 0$, and do not impose the NM constraint. If assumption (30) holds, the solution of this problem is $c_t \equiv y_t$ for all $t \leq T$ and $a_t \equiv a_t^*$ for all $t \leq L$, with probability 1.

3.2 Continuous Time

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

$$U = \begin{cases} E \left[\int_0^T e^{-\rho t} \frac{\left(c_t e^{-g(a_t)}\right)^{1-\gamma} - 1}{1-\gamma} dt \right] & \text{if } \gamma \neq 1 \\ E \left[\int_0^T e^{-\rho t} (\ln c_t - g(a_t)) dt \right] & \text{if } \gamma = 1. \end{cases} \tag{32}$$

For now, we consider the log utility case; in a later draft we will extend this section to

general CRRA utility functions. The firm's returns evolve according to:

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

where Z_t is a Brownian motion, and the volatility process σ_t is deterministic. We normalize $R_0 = 0$ and the risk premium to zero, i.e. the expected rate of return on the stock is r_f in each period.

Theorem 4 *(Optimal contract, continuous time, log utility). Let σ_t denote the stock volatility. The optimal contract pays the CEO c_t at each instant, where c_t satisfies:*

$$\ln c_t = \int_0^t \theta_s dr_s + \kappa_t, \tag{33}$$

where θ_s and κ_t are deterministic functions. Without the NM constraint:

$$\theta_t = \begin{cases} \frac{e^{-\rho t} g'(a_t^*)}{\int_t^T e^{-\rho \tau} d\tau} & \text{for } t \le L \\ 0 & \text{for } t > L. \end{cases} \tag{34}$$

With the NM constraint:

$$\theta_t = \begin{cases} \frac{\Theta}{\int_t^T e^{-\rho \tau} d\tau} & \text{for } t \le L + M \\ 0 & \text{for } t > L + M, \end{cases} \tag{35}$$

where $\Theta = \sup_{0 \le s \le L} \left(e^{-\rho s} g'(a_s^*) \right)$. Let $\zeta = -1$ denote the case if private savings are ruled out (and so the PS constraint is not imposed), and $\zeta = 1$ if they are allowed (and so the PS constraint is imposed). The value of κ_t is:

$$\kappa_t = (r_f - \rho) t - \int_0^t \theta_s E\left[dr_s \right] + \zeta \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds + \underline{\kappa}, \tag{36}$$

where $\underline{\kappa}$ ensures that the agent is at his reservation utility:

$$\underline{u} = \int_0^T e^{-\rho t} \left(\underline{\kappa} + (r_f - \rho) t + \zeta \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds - g(a_t^*) \right).$$

Proof (Heuristic). The Appendix presents a formal proof. Here, we provide a heuristic proof that conveys the "essence" of the economic argument. We used the techniques introduced

in Sannikov (2008). Consider the action / manipulation policy $(a, m) = (a_t, m_t)_{t \geq 0}$. By the martingale representation theorem, the agent's utility $U = \int_0^T e^{-\rho t} (\ln c_t - g(a_t^*) dt)$ can be written:

$$U = U_0 + \int_0^T Y_t (dr_t - a_t^* dt) \tag{37}$$

for a constant U_0 and a process Y_t adapted to the filtration induced by $(r_t)_{t \geq 0}$. The IC constraint (7) becomes:

$$IC_a: Y_t \geq e^{-\rho t} g'(a_t^*), \text{ for } t \leq L. \tag{38}$$

If there is no NM constraint, then the cost-minimizing incentive scheme entails the minimum sensitivity, so $Y_t = e^{-\rho t} g'(a_t^*)$ for $t \leq L$, and $Y_t = 0$ for $t > L$.

The NM constraint is:

$$Y_t = E_t[Y_{t+i}] \text{ for } 0 \leq t \leq L, 0 \leq i \leq M. \tag{39}$$

The cost-minimizing incentive scheme entails minimal sensitivities Y_t, subject to (38) and (39). The solution is $Y_t = \sup_{s \leq L} (\beta_s g'(a_s^*))$ for $t \leq L + M$, and $Y_t = 0$ for $t > L + M$.

The above considers a contract in terms of utility. We now translate it into a contract in terms of consumption. Using again the martingale representation theorem, we can write:

$$\ln c_t = \kappa_t + \int_0^t \xi_{st} \sigma_s dz_s,$$

where $(\xi_{st})_{s \leq t \leq T}$ is an $\mathcal{F}_t-$adapted process, and κ_t is a deterministic number. ξ_{st} is the sensitivity of time-t consumption to a past shock, dr_s. Intuitively, to minimize the cost of the contract while keeping the agent's utility fixed, the principal wishes to smooth consumption. Therefore, for a given shock at time s, the sensitivity of all future consumptions should be the same. We thus have, for all $t \geq s$, $\xi_{st} = \theta_s$, for some value θ_s. Hence, we can calculate:

$$U = \int_0^T e^{-\rho t} (\ln c_t - g(a_t^*)) = K + \int_0^T e^{-\rho t} \left(\int_0^t \theta_s \sigma_s dz_s \right) dt = K + \int_0^T \left(\int_s^T e^{-\rho t} dt \right) \theta_s \sigma_s dz_s.$$

We require that $U = U_0 + \int_0^T Y_s \sigma_s dz_s$. This implies $\theta_s = \frac{Y_s}{\int_s^T \beta_\tau d\tau}$, as given in Theorem 4.

The expression for κ_t comes from the PS constraint (9), or, if it is not imposed, the IEE (10). ∎

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

3.3 The Maximum Effort Principle

The focus of this paper is to derive the optimal contract to implement a given path of effort levels. EG shows that, if the firm is sufficiently large, the maximum effort level is optimal. While they considered a discrete time, one-period setting, we extend this maximum effort principle to continuous time. The analysis is still in progress; this section contains the results obtained thus far.

We consider a continuous-time model where earnings follow $dD_t/D_t = (a_t + k)\, dt + \sigma dZ_t$. Let $S = D_0/r$ denote the baseline firm size.

Theorem 5 *Fix $\underline{u}$. For any $\varepsilon > 0$ there exists S_* large enough such, if $S > S_*$, the principal's profit from the contract in Theorem 4 differs from his profit from the optimal contract by at most ε.*

Hence, we show that contract requiring maximum effort is optimal, within an ε. The proof is will be made available soon. We suspect that an analogous, and stronger result, might be available in discrete time. We are currently researching this issue.

4 Conclusion

This paper studies optimal CEO compensation in a dynamic setting in which the CEO consumes in each period, can privately save, and may manipulate current earnings at the expense of future profits. The optimal contract involves consumption smoothing, where effort is rewarded in all future periods, and the relevant measure of incentives is the percentage change in pay for a percentage change in firm value. This required sensitivity is constant over time in an infinite horizon model where manipulation is impossible. If the horizon is finite, the contract's slope rises over time since, as the CEO approaches retirement, he has fewer periods over which to be rewarded for effort. A rising slope also arises if the contract needs to prevent manipulation. This is to offset the fact that the cost of manipulation is suffered only

in the future and thus has a discounted effect on the CEO's utility. Deterring manipulation also requires the CEO to remain sensitive to the firm's stock price after retirement. While the possibility of manipulation affects the sensitivity of pay to firm performance, the option to privately save impacts the time trend in total pay. Specifically, it augments the rise in compensation over time, to deter the CEO from saving to finance future consumption.

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

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

A Proofs

A.1 Proof of Theorem 1

This is a direct corollary of Theorem 2.

A.2 Proof of Theorem 2

We first analyze the case without the NM constraints. We consider the NM constraints at the end of the proof.

Case $t > L$. For $t > L$, r_t is independent of the CEO's actions. Since the CEO is also strictly risk averse, the efficient contract will have c_t for $t > L$ depend only on $r_1, ..., r_L$. Therefore either the PS constraint (9) or IEE (10) immediately give

$$\ln c_t(r_1, ..., r_t) = \ln c_L(r_1, ..., r_L) + \kappa_t, \tag{40}$$

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

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

$$0 \in \arg\max_{\varepsilon \leq 0} U(r_1, ..., r_{L-1}, a_L^* + \eta_L + \varepsilon). \tag{41}$$

Since g is differentiable, this yields (7) (see e.g. EG, Lemma 6), i.e.

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

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

and so

$$\frac{d}{d\varepsilon_-} \ln c_L(r_1, ..., a_L^* + \eta_L + \varepsilon) \geq \frac{B_t g'(a_t^*)}{\sum_{s=L}^{T} B_t e^{(1-\gamma)k_t}} := \theta_L. \tag{42}$$

We now show that indeed (42) holds with equality. If a_L^* is interior, this follows immediately from the IC requirement analogous to (41) but for $\varepsilon \geq 0$. Here we prove this fact for

34

the case $a_L^* = \bar{a}_L$. The result is intuitive, as a binding constraint will minimize the variability in the agent's pay and constitute efficient risk-sharing. First, condition (42) implies that for any $r' \geq r$ (see e.g. EG, Lemma 4)

$$\ln c_L (r_1, ...r_{L-1}, r') - \ln c_L (r_1, ...r_{L-1}, r) \geq \theta_L * (r' - r). \tag{43}$$

Consider now the contract $\{c_t^0\}_{t \leq T}$ that coincides with $\{c_t\}_{t \leq T}$ for $t < L$, $\ln c_t^0 = \ln c_L^0 + \kappa_t$ for $t > L$ and κ_t as in (40), and such that $c_L^0(r_1, ..., r_L) = e^{B(r_1,...,r_{L-1})+\theta_L r_L}$, where $B(r_1, ..., r_{L-1})$ is chosen to satisfy

$$E_{L-1}\left[\frac{(c_L^0)^{1-\gamma}(r_1, ..., r_L)}{1-\gamma}\right] = E_{L-1}\left[\frac{(c_L)^{1-\gamma}(r_1, ..., r_L)}{1-\gamma}\right]. \tag{44}$$

Note that the condition (43) guarantees that the random variable $\ln c_L (r_1, ...r_{L-1}, \widetilde{r}_L)$ is weakly more dispersed than $\ln c_L^0 (r_1, ...r_{L-1}, \widetilde{r}_L)$. It also follows from the IC that both $\ln c_L (r_1, ...r_{L-1}, \cdot)$ and $\ln c_L^0 (r_1, ...r_{L-1}, \cdot)$ are weakly increasing. Those facts together with (44) imply that for the convex function ψ and increasing function ξ, where $\psi^{-1}(x) = \frac{x^{1-\gamma}}{1-\gamma}$ and $\xi(x) = \frac{e^{(1-\gamma)x}}{1-\gamma}$, we have (see EG, Lemmas 1 and 2):

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

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

Integrating out (42) that holds with equality, the optimal contract c is given by:

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

for some function B.

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

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

35

for some function B as well as θ_s and $\chi_{t'}$ as in the Theorem. The IEE yields

$$\frac{1}{c_t^{-\gamma}} = e^{-r_f} \frac{B_t}{B_{t+1}} E_t \left[\frac{1}{c_{t+1}^{-\gamma}} \right] = E_t \left[e^{\gamma(\theta_{t+1}+\chi_{t+1})r_{t+1}} \right] e^{-r_f + \ln B_t - \ln B_{t+1} + \gamma B(r_1,...,r_t) + \gamma\kappa_{t+1}}. \quad (45)$$

whereas the PS constraint yields

$$c_t^{-\gamma} = e^{r_f} \frac{B_{t+1}}{B_t} E_t \left[c_{t+1}^{-\gamma} \right] = E_t \left[e^{-\gamma(\theta_{t+1}+\chi_{t+1})r_{t+1}} \right] e^{-\gamma B(r_1,...,r_t) + r_f - \gamma\kappa_{t+1} + \ln B_{t+1} - \ln B_t}. \quad (46)$$

In either case we therefore have

$$\ln c_t = B(r_1, ..., r_t) + \phi(\theta_{t+1} + \chi_{t+1})r_t + \kappa_t = B(r_1, ..., r_t) + \chi_t r_t + \kappa_t. \quad (47)$$

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

$$B_t c_t^{1-\gamma} \chi_t + \frac{d}{d\varepsilon_-} B\left(r_1, ...r_{t-1}, a_t^* + \eta_t + \varepsilon\right) \sum_{s=t}^T B_s E_t \left(c_s^{1-\gamma}\right) \geq B_t c_t^{1-\gamma} g'(a_t^*), \quad (48)$$

$$B_t c_t^{1-\gamma} \chi_t + \frac{d}{d\varepsilon_-} B\left(r_1, ...r_{t-1}, a_t^* + \eta_t + \varepsilon\right) c_t^{1-\gamma} \sum_{s=t}^T B_s e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} \times$$

$$\times E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^s (\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^s \theta_m a_m^* + \chi_s a_s^*\right]}\right) \geq B_t c_t^{1-\gamma} g'(a_t^*),$$

$$\frac{d}{d\varepsilon_-} B\left(r_1, ...r_{t-1}, a_t^* + \eta_t + \varepsilon\right) \geq$$

$$\geq \frac{B_t\left(g'(a_t^*) - \chi_t\right)}{\sum_{s=t}^T B_s e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^s (\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^s \theta_m a_m^* + \chi_s a_s^*\right]}\right)} := \theta_t.$$

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

$$E_t \left[c_s^{1-\gamma}\right] = c_t^{1-\gamma} e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E_t \left[e^{(1-\gamma)\left(\sum_{m=t+1}^s \theta_m(a_m^* + \eta_m) + \chi_S(a_s^* + \eta_s) - \chi_t \eta_t\right)}\right] =$$

$$= c_t^{1-\gamma} e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E_t \left[e^{(1-\gamma)\left(\sum_{m=t+1}^s \theta_m a_m^* + \chi_s a_s^* + \sum_{m=t+1}^s \theta_m \eta_m + \chi_s \eta_s - \chi_t \eta_t\right)}\right] =$$

$$= c_t^{1-\gamma} e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E_t \left[e^{(1-\gamma)\left(\sum_{m=t+1}^s \theta_m a_m^* + \chi_s a_s^* + \sum_{m=t+1}^s (\theta_m + \chi_m)\varepsilon_m\right)}\right].$$

One can inductively show that for any $t \leq L$, $\theta_t + \chi_t \leq g'(a_t^*)$ and, using (23) and (48), $\theta_t \geq 0$. Therefore, proceeding analogously as in the proof of the case $t = L$ we can establish that indeed (48) holds with equality. Integrating out this equality and using (17), we establish

that for $t' \geq t$

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

where θ_t and χ_t are as required.

We now determine the values of the constants κ_t. When the PS constraint is not imposed, we use the IEE (10). First, there exists a value $e^{\underline{\kappa}}$ such that $e^{\underline{\kappa}} = \frac{e^{-r_f t} E[c_t^\gamma]}{B_t}$ for $t \leq T$. for all t. This yields, for all t,

$$\gamma \kappa_t = \underline{\kappa} + r_f t + \ln B_t - \ln E\left[e^{\gamma\left(\sum_{s=1}^{t} \theta_s r_s + \chi_t r_t\right)}\right],$$

where

$$\ln E\left[e^{\gamma\left(\sum_{s=1}^{t} \theta_s r_s + \chi_t r_t\right)}\right] = \ln E\left[E_{t-1}\left[e^{\gamma\left(\sum_{s=1}^{t} \theta_s r_s + \chi_t r_t\right)}\right]\right] =$$

$$\ln E\left[e^{\gamma\left(\sum_{s=1}^{t-1} \theta_s r_s + (\theta_t + \chi_t)\phi r_{t-1}\right)} E_{t-1}\left[e^{\gamma(\theta_t + \chi_t)(a_t^* - \phi a_{t-1}^* + \varepsilon_t)}\right]\right] = \sum_{s=1}^{t} \ln E\left[e^{\gamma(\theta_s + \chi_s)(a_s^* - \phi a_{s-1}^* + \varepsilon_s)}\right],$$

i.e. (14) with $\zeta = -1$. When the PS constraint is imposed, we use (9). There exists a value $e^{-\underline{\kappa}}$ such that $e^{-\underline{\kappa}} = e^{r_f t} B_t E\left[c_t^{-\gamma}\right]$ for all t. This yields, just as above, (14) with $\zeta = 1$. Finally, constant $\underline{\kappa}$ is chosen to let the agent be at his reservation utility.

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

$$\ln c_t = \sum_{s=1}^{t} \theta'_s r_s + \chi'_t r_t + \kappa'_t,$$

with $\chi'_t = \theta'_t = 0$ for $t > L + M$, $\chi'_t = \phi(\chi'_{t+1} + \theta'_{t+1})$ and κ'_t as in the Theorem, whereas θ'_t are the lowest numbers such that the IC constraint is satisfied, i.e.

$$\theta'_t \geq \frac{B_t\left(g'(a_t^*) - \chi'_t\right)}{\sum_{s=t}^{T} B_s e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^{s}(\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^{s} \theta_m a_m^* + \chi_s a_s^*\right]}\right)}, \quad \text{for } 0 \leq t \leq L,$$

$$(49)$$

and such that the NM constraint holds:

$$E_t\left[\frac{\partial U}{\partial r_t}\right] = E_t\left[\frac{\partial U}{\partial r_{t+i}}\right], \text{ for } 0 \le t \le L,\, 0 \le i \le M. \tag{50}$$

Now, if we let

$$\theta'_{L+i} = \frac{D_i}{\sum_{s=L+i}^{T} B_s e^{(1-\gamma)(\kappa'_s - \kappa'_{L+M})}},$$

for some constants D_i, $i \le M$, the (50) is equivalent to

$$B_t c_t^{1-\gamma} \chi'_t + \theta'_t c_t^{1-\gamma} \sum_{s=t}^{T} B_s e^{(1-\gamma)(\kappa_s - \kappa_t - \chi_t a_t^*)} E\left(e^{(1-\gamma)\left[\sum_{m=t+1}^{s}(\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^{s}\theta_m a_m^* + \chi_s a_s^*\right]}\right) =$$

$$= E_t\left[c_{L+i}^{1-\gamma} \times D_i\right] = c_t^{1-\gamma} e^{(1-\gamma)(\kappa_{L+i} - \kappa_t - \chi_t a_t^*)} E_t\left[e^{(1-\gamma)\sum_{m=t+1}^{L+i}(\theta_m + \chi_m)\varepsilon_m + \sum_{m=t+1}^{L}\theta_m a_m^*}\right] \times D_i,$$

for $0 \le t \le L$, $i \le M$. This yields the desired formulas for θ'_t, $t \le L + M$, with $D = D_M$.

A.3   Proof of Theorem 3

We begin with the following lemma.

Lemma 1 *(Concavity of Present Values) Let*

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

where all j_s and h_t are twice differentiable functions. Suppose that[7]:

$$\sup\left(4\min\{T, \frac{e^{(\sup h_t - \inf h_t)/2}}{1 - \min\{1, \delta\}}\}\max\{h_s'^2, j_s'^2\} + j_s''\right) < 0, \tag{51}$$

$$\sup\left(2h_t' \max\{h_t', j_t'\} + h_t''\right) < 0,$$

for $\delta = e^{\frac{\sup j_t}{2}}$. Then the function I is concave.

Loosely speaking, the Lemma says that, if j_s and h_t are sufficiently concave functions, then the "PV of income function" $I(b_t)$ associated with them is also concave. This is non-

---

[7]All the suprema and infima are taken with respect to both the arguments and the time indices.

trivial to prove when T is infinite: for sufficiently large T, the function $\exp(Tj(b))$ will be convex. It is discounting (expressed by $\delta < 1$) that allows the income function to be concave.

Proof Let

$$P_s((b_t)_{t \leq L}) = e^{\sum_{n=1}^{s} j_n(b_n) + h_s(b_s)},$$

$$S_s((b_t)_{t \leq L}) = \sum_{n=s}^{T} e^{\sum_{m=1}^{n} j_m(b_m) + h_n(b_n)} = \sum_{n=s}^{T} P_n((b_t)_{t \leq L}),$$

for any $s \leq T$. Fix for the rest of the proof an argument sequence $(b_t)_{t \leq L}$ (we will evaluate all the functions at this sequence, and consequently economize on notation writing e.g. S_r instead of $S_r((b_t)_{t \leq L})$ or h_n instead of $h_n(b_n)$).

Step 1: Derivatives. For unit vectors e_r and e_s in the direction r and s, $r \geq s$, consider the derivatives of the function I:

$$\frac{\partial I}{\partial e_s} = j_s' S_s + h_s' P_s,$$

$$\frac{\partial^2 I}{\partial e_r \partial e_s} = j_r' j_s' S_r + h_r' j_s' P_r + \mathbf{1}_{r=s} \left(h_s'(h_s' + j_s') P_s + j_s'' S_s + h_s'' P_s \right).$$

For a fixed vector $\mathbf{y} = (y_t)_{t \leq T}$ the second derivative in the direction $\mathbf{y} = (y_t)_{t \leq T}$ is:

$$\frac{\partial^2 I}{\partial \mathbf{y} \partial \mathbf{y}} = \sum_{s=1}^{T} y_s \sum_{r=1}^{T} y_r \frac{\partial^2 I}{\partial e_s \partial e_r} =$$

$$= \sum_{s=1}^{T} y_s^2 \left(h_s'(h_s' + j_s') P_s + j_s'' S_s + h_s'' P_s \right) + \sum_{s=1}^{T} j_s' y_s^2 (j_s' S_s + h_s' P_s) + 2 \sum_{s,r>s} j_s' y_s y_r (j_r' S_r + h_r' P_r).$$

Step 2: Bounding the S_r sums. For any $s \leq T$ and $q \leq T - s$ we have $S_{s+q} \leq S_s$. Moreover,

$$S_{s+q} = \sum_{t=s+q}^{T} e^{\sum_{n=1}^{t} j_n + h_t} \leq e^{\sup h_t} \sum_{t=s+q}^{T} e^{\sum_{n=1}^{t} j_n} \leq$$

$$\leq e^{\sup h_t} e^{q \sup j_t} \sum_{t=s}^{T-q} e^{\sum_{n=1}^{t} j_n} \leq e^{q \sup j_t} S_s e^{\sup h_t - \inf h_t}.$$

It follows that for any $\psi \in (\sup j_t, 0) \cup \{0\}$ we have:

$$\sum_{r \geq s} S_r e^{-\psi(r-s)} \leq C_1 S_s,$$

$$\sum_{s,r \geq s} S_r y_r^2 e^{\psi(r-s)} = \sum_r y_r^2 S_r \sum_{s \leq r} e^{\psi(r-s)} \leq C_2 \sum_s S_s y_s^2,$$

where

$$C_1 = C_2 = T, \text{ for } \psi = 0, \tag{52}$$

$$C_1 = \frac{e^{\sup h_t - \inf h_t}}{1 - e^{\sup j_t - \psi}}, \quad C_2 = \frac{1}{1 - e^{\psi}} \text{ otherwise.}$$

Step 3: Bounding the derivatives. For any $\psi \in (\sup j, 0) \cup \{0\}$ and any vector $\mathbf{y} = (y_t)_{t \leq T}$, we have:

$$\sum_{s,r \geq s} y_s y_r S_r = \sum_s y_s \sum_{r \geq s} \sqrt{S_r} y_r e^{\frac{\psi}{2}(r-s)} \sqrt{S_r} e^{-\frac{\psi}{2}(r-s)} \leq \sum_s y_s \left(\sum_{r \geq s} S_r y_r^2 e^{\psi(r-s)} \right)^{1/2} \left(\sum_{r \geq s} S_r e^{-\psi(r-s)} \right)^{1/2} \leq$$

$$\leq \sqrt{C_1} \sum_s y_s \sqrt{S_s} \left(\sum_{r \geq s} S_r y_r^2 e^{\psi(r-s)} \right)^{1/2} \leq \sqrt{C_1} \left(\sum_s y_s^2 S_s \right)^{1/2} \left(\sum_s \left(\sum_{r \geq s} S_r y_r^2 e^{\psi(r-s)} \right) \right)^{1/2} \leq$$

$$\leq \sqrt{C_1 C_2} \left(\sum_s y_s^2 S_s \right)^{1/2} \left(\sum_s S_s y_s^2 \right)^{1/2} = \sqrt{C_1 C_2} \sum_s y_s^2 S_s,$$

where the first and third inequalities follow from the Cauchy-Schwartz inequality, and C_1 and C_2 are as in (52). Therefore:

$$\frac{\partial^2 I}{\partial \mathbf{y} \partial \mathbf{y}} = \sum_{s=1}^T y_s^2 \left(h_s'(h_s' + j_s') P_s + j_s'' S_s + h_s'' P_s \right) + \sum_{s=1}^T j_s' y_s^2 (j_s' S_s + h_s' P_s) + 2 \sum_{s,r > s} j_s' y_s y_r (j_r' S_r + h_r' P_r) \leq$$

$$\leq \sum_{s=1}^T y_s^2 \left(2 h_s' \max\{h_s', j_s'\} P_s + j_s'' S_s + h_s'' P_s \right) + 2 \sum_{s,r \geq s} y_s \max\{j_s', h_s'\} y_r 2 \max\{h_r', j_r'\} S_r \leq$$

$$\leq \sum_{s=1}^T y_s^2 P_s \left(2 h_s' \max\{h_s', j_s'\} + h_s'' \right) + \sum_{s=1}^T y_s^2 S_s \left(4\sqrt{C_1 C_2} \max\{h_s'^2, j_s'^2\} + j_s'' \right).$$

Letting $\psi = 0$ or $\psi = \frac{\sup j_t}{2}$ in case $\sup j_t < 0$ proves the Lemma. ∎

We now move to the proof of the Theorem itself. For this proof it is helpful to introduce some new notation. Let $a_t = f(x_t)$, where

$$
x_t = \begin{cases} e^{-g(a_t)\frac{\gamma}{1-\gamma}} & \text{if } \gamma < 1 \\ -g(a_t) & \text{if } \gamma = 1 \\ e^{-g(a_t)} & \text{if } \gamma > 1. \end{cases}
$$

x_t measures the agent's leisure, and f is the "production function" given leisure, which is decreasing and concave.

Case $\gamma = 1$. Consider $c_t^*(\eta) = \exp\left(\sum_{n=1}^{t} \theta_n(\eta_n + f(x_n^*)) + \chi_t(\eta_t + f(x_t^*)) + \kappa_t\right)$, the savings-free consumption for the recommended path of actions on the path of noises $\eta = (\eta_t)_{t \leq T}$ (where $x_t^* = -g(a_t^*)$). For any path of noises $\eta = (\eta_t)_{t \leq T}$ we introduce the "upper linearization" utility function $\widehat{U}_\eta$ defined by:

$$
\widehat{U}_\eta\left((c_t)_{t \leq T}, (x_t)_{t \leq L}\right) = \sum_{t=1}^{T} e^{-\rho t}(\ln c_t^*(\eta) - 1) + \sum_{t=1}^{T} e^{-\rho t}\left(\frac{c_t}{c_t^*(\eta)}\right) + \sum_{t=1}^{L} e^{-\rho t} x_t.
$$

By construction, $\widehat{U}_\eta = U + \sum_{t=1}^{T}(c_t - c_t^*(\eta))\frac{\partial U\left((c_t^*(\eta))_{t \leq T}, (x_t^*)_{t \leq L}\right)}{\partial c_t} + \sum_{t=1}^{L}(x_t - x_t^*)\frac{\partial U\left((c_t^*(\eta))_{t \leq T}, (x_t^*)_{t \leq L}\right)}{\partial x_t}$.
Since U is concave, we have the key property:

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

Hence, to show that there are no profitable deviations for EU, it is sufficient to show that there are no profitable deviations for $E\widehat{U}_\eta$. Since $e^{(r_f - \rho)t}/c_t^*(\eta)$ is a martingale, the agent is indifferent at which time he consumes income y_t, so we evaluate $E\widehat{U}_\eta$ for $c_t \equiv y_t$. The utility

on any path η is then, with a slight abuse of notation[8]:

$$\widehat{U}_\eta((y_t)_{t\leq T}, (x_t)_{t\leq L}) = \sum_{t=1}^T e^{-\rho t}\left(\ln c_t^*(\eta) - 1\right) + \sum_{t=1}^T e^{-\rho t}\left(\frac{y_t}{c_t^*(\eta)}\right) + \sum_{t=1}^L e^{-\rho t}x_t =$$

$$= \sum_{t=1}^T e^{-\rho t}\left(\ln c_t^*(\eta) - 1\right) + \sum_{t=1}^T e^{\sum_{n=1}^t \theta_n(f(x_n)-a_n^*)+\chi_t(f(x_t)-a_t^*)-\rho t} + \sum_{t=1}^L e^{-\rho t}x_t.$$

We now use Lemma 1 applied to $I((x_t)_{t\leq L}) = \sum_{t=1}^T e^{\sum_{n=1}^t \theta_n(f(x_n)-a_n^*)+\chi_t(f(x_t)-a_t^*)-\rho t}$ with $j_s(x_s) = \theta_s(f(x_s) - a_s^*) - \rho$ and $h_s(x_s) = \chi_s(f(x_s) - a_s^*)$. Since

$$f'(x_s) = \frac{-1}{g'(f(x_s))}, \quad f''(x_s) = \frac{g''(f(x_s))}{g'^2(f(x_s))},$$

and $\theta_s \leq g'(a_s^*)$, condition CONV implies (51).

It follows that $\widehat{U}_\eta((y_t)_{t\leq T}, \cdot)$ is concave, for every η. Since due to Theorem 2 $\widehat{U}_\eta((y_t)_{t\leq T}, \cdot)$ also satisfies the FOC at $(x_t^*)_{t\leq L}$, it is maximized at $(x_t^*)_{t\leq L}$, for every η. Therefore $E\widehat{U}_\eta$ is maximized by (the processes) $c_t \equiv y_t$ and $x_t \equiv x_t^*$, proving the Theorem for this case.

Case $\gamma \neq 1, \alpha = 0$. As before, let $c_t^*(\eta) = \exp\left(\sum_{n=1}^t \theta_n(\eta_n + f(x_t^*)) + \kappa_t\right)$ be the savings-free consumption for the recommended path of actions on the path of noises $\eta = (\eta_t)_{t\leq T}$ (where $x_t^* = e^{-g(a_t^*)}$ for $\gamma > 1$, and $x_t^* = e^{-g(a_t^*)\frac{\gamma}{1-\gamma}}$ for $\gamma < 1$). The variables x_t have been defined in such a way that the U function is concave. The "upper linearization" utility function $\widehat{U}_\eta = U + \sum_{t=1}^T (c_t - c_t^*(\eta))\frac{\partial U\left((c_t^*(\eta))_{t\leq T},(x_t^*)_{t\leq L}\right)}{\partial c_t} + \sum_{t=1}^L (x_t - x_t^*)\frac{\partial U\left((c_t^*(\eta))_{t\leq T},(x_t^*)_{t\leq L}\right)}{\partial x_t}$ is now:

$$\widehat{U}_\eta\left((c_t)_{t\leq T}, (x_t)_{t\leq L}\right) = \frac{\sum_{t=1}^T e^{-\rho t}(c_t^*(\eta)x_t^*)^{1-\gamma}\left(\frac{1}{1-\gamma} - 2\right) + }{\sum_{t=1}^T e^{-\rho t}x_t^{*1-\gamma}\left(\frac{c_t}{c_t^{*\gamma}(\eta)}\right) + \sum_{t=1}^L e^{-\rho t}(c_t^*(\eta))^{1-\gamma}\left(\frac{x_t}{x_t^{*\gamma}}\right)}.$$

---

[8]The abuse of notation arises since x_t is not defined for $t > L$. This is irrelevant, however, as $\theta_t = \chi_t = 0$ for $t > L$;

Since, as before, agent finds it optimal to consume his own income, it is enough to show that there are no profitable deviations from $E\widehat{U}_\eta((y_t)_{t\leq T}, \cdot)$:

$$E\widehat{U}_\eta((y_t)_{t\leq T}, (x_t)_{t\leq L}) =$$

$$E\left[\begin{array}{c} \sum_{t=1}^T e^{-\rho t}(c_t^*(\eta)x_t^*)^{1-\gamma}\left(\frac{1}{1-\gamma}-2\right) + \sum_{t=1}^L e^{-\rho t}\left(c_t^*(\eta)\right)^{1-\gamma}\left(\frac{x_t}{x_t^{*\gamma}}\right) + \\ + \sum_{t=1}^T e^{-\rho t}x_t^{*1-\gamma}e^{(1-\gamma)\left[\kappa_t+\sum_{n=1}^t \theta_n\varepsilon_n\right]+\sum_{n=1}^t \theta_n(f(x_n)-\gamma a_n^*)} \end{array}\right] =$$

$$E\left[\begin{array}{c} \sum_{t=1}^T e^{-\rho t}(c_t^*(\eta)x_t^*)^{1-\gamma}\left(\frac{1}{1-\gamma}-2\right) + \sum_{t=1}^L e^{-\rho t}\left(c_t^*(\eta)\right)^{1-\gamma}\left(\frac{x_t}{x_t^{*\gamma}}\right) + \\ + \sum_{t=1}^T e^{(1-\gamma)\sum_{n=1}^t (\theta_n\varepsilon_n-\ln E(e^{\theta_n e_n}))}e^{-\rho t+(1-\gamma)\left[\kappa_t-g(a_t^*)+\sum_{n=1}^t \ln E(e^{\theta_n e_n})\right]+\sum_{n=1}^t \theta_n(f(x_n)-\gamma a_n^*)} \end{array}\right].$$

Since $M_t = e^{(1-\gamma)\sum_{n=1}^t (\theta_n\varepsilon_n-\ln E(e^{\theta_n e_n}))}$ is a positive martingale, we have that

$$E\left[\sum_{t=1}^T e^{-\rho t}\left(c_t^*(\eta)\right)^{1-\gamma}\left(\frac{x_t}{x_t^{*\gamma}}\right) + \sum_{t=1}^T M_t e^{-\rho t+(1-\gamma)\left[\kappa_t-g(a_t^*)+\sum_{n=1}^t \ln E(e^{\theta_n e_n})\right]+\sum_{n=1}^t \theta_n(f(x_n)-\gamma a_n^*)}\right] =$$

$$= E^Q\left[\sum_{t=1}^T e^{-\rho t}\left(c_t^*(\eta)\right)^{1-\gamma}\left(\frac{x_t}{M_t x_t^{*\gamma}}\right) + \sum_{t=1}^T e^{-\rho t+(1-\gamma)\left[\kappa_t-g(a_t^*)+\sum_{n=1}^t \ln E(e^{\theta_n e_n})\right]+\sum_{n=1}^t \theta_n(f(x_n)-\gamma a_n^*)}\right] =$$

$$=: E^Q\left[U_\eta^{-Q}((x_t)_{t\leq L})\right],$$

where measure Q is defined by $E^Q[A] = E[M_T A]/M_0$ for all events A, i.e. Q has Radon-Nikodym derivative w.r.t. P equal to M_T/M_0. We now apply Lemma 1 to $I((x_t)_{t\leq L}) = e^{-\rho t+(1-\gamma)\left[\kappa_t-g(a_t^*)+\sum_{n=1}^t \ln E(e^{\theta_n e_n})\right]+\sum_{n=1}^t \theta_n(f(x_n)-\gamma a_t^*)}$ with

$$j_s(x_s) = \theta_s(f(x_s) - \gamma a_s^*) + (1-\gamma)\left[(\kappa_s - g(a_s^*)) - (\kappa_{s-1} - g(a_{s-1}^*)) + \ln E(e^{\theta_s e_s})\right] - \rho,$$

$$h_s(x_s) \equiv 0.$$

Since

$$f'(x_s) = -C\frac{1}{x_s g'(f(x_s))},$$

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

with $C = 1$ for $\gamma > 1$ and $C = \frac{\gamma}{1-\gamma}$ for $\gamma < 1$, and since also $\theta_s \leq g'(a_s^*)$, we have

$$4\min\{T, \frac{e^{(\sup h_t - \inf h_t)/2}}{1-\delta}\}\max\{h_s'^2, j_s'^2\} + j_s'' =$$

$$4\min\{T, \frac{1}{1-\min\{\delta,1\}}\}\frac{\theta_t^2 C^2}{x_t^2 g'^2(f(x_t))} + \frac{\theta_t}{g'^2(f(x_t))}\left(C\frac{g'(f(x_t))}{x_t^2} - C^2\frac{g''(f(x_t))}{x_t^2 g'(f(x_t))}\right) =$$

$$\frac{\theta_t C^2}{x_t^2 g'^3(f(x_t))}\left[4\min\{T, \frac{1}{1-\min\{\delta,1\}}\}g'(f(x_t))\theta_t + \frac{g'^2(f(x_t))}{C} - g''(f(x_t))\right] \leq$$

$$\frac{\theta_t C^2}{x_t^2 g'^3(f(x_t))}\left[g'^2(f(x_t))\left(4\min\{T, \frac{1}{1-\min\{\delta,1\}}\} + \frac{1}{C}\right) - g''(f(x_t))\right],$$

and so condition CONV implies (51).

Therefore, $U_\eta^{-Q}(\cdot)$ is concave for every η, and consequently $E\widehat{U}_\eta((y_t)_{t\leq T}, \cdot) = E^Q\left[U_\eta^{-Q}(\cdot)\right] + E\left[\sum_{t=1}^{T} e^{-\rho t}(c_t^*(\eta)x_t^*)^{1-\gamma}\left(\frac{1}{1-\gamma} - 2\right)\right]$ is concave in (the process) $(x_t)_{t\leq L}$. Since due to Theorem 2 $E\widehat{U}_\eta((y_t)_{t\leq T}, \cdot)$ also satisfies the FOC at (the process) $(x_t^*)_{t\leq L}$, it is maximized at $(x_t^*)_{t\leq L}$. This establishes the Theorem.

A.4 Proof of Theorem 4

The arguments in the heuristic proof yield the value of Y_t. We now rigorously derive the associated consumption path. This proof is currently in progress and will be formalized and sharpened in a future draft.

Case where there is no PS constraint

Using again the martingale representation theorem, we can write:

$$\ln c_t = \kappa_t' + \int_0^t \xi_{st}\sigma_s dz_s, \tag{53}$$

where $(\xi_{st})_{s\leq t\leq T}$ is an $\mathcal{F}_t$−adapted process, and κ_t' is a deterministic constant. Therefore, with $U = \int e^{-\rho t}\ln(c_t - g(a_t^*))$, we have:

$$U = K + \int_0^T e^{-\rho t}\left(\int_0^t \xi_{st}\sigma_s dz_s\right)dt = \int_0^T\left(\int_s^T e^{-\rho t}\xi_{st}dt\right)\sigma_s dz_s.$$

We thus have, a.s.:

$$Y_s = \int_s^T e^{-\rho t}\xi_{st}dt. \tag{54}$$

44

This expression states that the sensitivity Y_s of total utility to a shock dz_s is the present value of the sensitivities ξ_{st} of future instantaneous utilities.

The principal wishes to minimize expected cost $E\left[\int_0^T e^{-r_f t} c_t dt\right]$ subject to (53), (54) and the participation constraint, $\int e^{-\rho t}\kappa_t' dt \geq \underline{u}$. We form the Lagrangian

$$\mathcal{L} = -E\left[\int_0^T e^{-r_f t} e^{\kappa_t' + \int_0^t \xi_{st}\sigma_s dz_s}\right] + \mu \int e^{-\rho t}\kappa_t' dt + \int \pi_s \left(\int_s^T e^{-\rho t}\xi_{st}dt - Y_s\right) ds$$

and minimize over κ_t' and ξ_{st}.

$$\frac{\partial \mathcal{L}}{\partial \kappa_t} = -E\left[e^{-r_f t}e^{\kappa_t' + \int_0^t \xi_{st}\sigma_s dz_s}\right] + \mu e^{-\rho t} \tag{55}$$

i.e., $E\left[e^{-r_f t}c_t\right] = \mu e^{-\rho t}$, the Inverse Euler equation. The optimization on ξ_{st} yields[9]

$$\frac{\partial \mathcal{L}}{\partial \xi_{st}} = -E\left[e^{-r_f t}e^{\kappa_t' + \int_0^t \xi_{st}\sigma_s dz_s}\xi_{st}\sigma_s^2\right] + \pi_s e^{-\rho t}, \tag{56}$$

and so we have $\xi_{st} = \pi_s/(\mu\sigma_s^2) \equiv \theta_s$. This means that the sensitivities of future log consumption to the shock dz_s are identical, which is intuitive given the desirability of consumption smoothing. Equation (54) gives $Y_s = \theta_s \int_s^T e^{-\rho t}dt$, hence $\theta_s = Y_s/\left(\int_s^T e^{-\rho t}dt\right)$. Plugging this into (55), we have:

$$\exp\left(-r_f t + \kappa_t' + \int_0^t \frac{\theta_s^2 \sigma_s^2}{2}ds\right) = \mu e^{-\rho t}.$$

and thus $\kappa_t' = \kappa^0 + (r_f - \rho)t - \int_0^t \frac{\theta_s^2 \sigma_s^2}{2}ds$ for some constant κ^0. Since $dr_s = \sigma_s dz_s + E\left[dr_s\right]$, we obtain, for $\kappa_t = \kappa_t' - \int_0^t \theta_s E\left[dr_s\right]$:

$$\ln c_t = \kappa^0 + (r_f - \rho)t - \int_0^t \frac{\theta_s^2 \sigma_s^2}{2}ds + \int_0^t \theta_s \sigma_s dz_s = \kappa^0 + (r_f - \rho)t - \int_0^t \frac{\theta_s^2 \sigma_s^2}{2}ds - \int_0^t \theta_s E\left[dr_s\right] + \int_0^t \theta_s dr_t$$

$$= \kappa_t + \int_0^t \theta_s dr_t,$$

---

[9]We use the fact that for a function F,

$$\frac{d}{dx_t}E\left[F\left(\int_0^t x_s\sigma_s dz_s\right)\right] = E\left[F''\left(\int_0^t x_s dz_s\right)x_t\sigma_t^2\right],$$

which can be verified via Ito's formula.

which gives the announced expression for c_t.

Case where there is a PS constraint

By (9), $e^{r_f t} e^{-\rho t}/c_t$ is a positive martingale. Therefore, there is an adapted process θ_s such that:

$$e^{r_f t} e^{-\rho t}/c_t = \exp\left(-\int_0^t \theta_s \sigma_s dz_s - \frac{\theta_s^2 \sigma_s^2}{2} ds\right)$$

Hence, we have, for some constant κ^0,

$$\ln c_t = \kappa^0 + r_f t + (r_f - \rho) t + \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds + \int_0^t \theta_s \sigma_s dz_s$$

$$= \kappa^0 + r_f t + (r_f - \rho) t + \int_0^t \frac{\theta_s^2 \sigma_s^2}{2} ds - \int_0^t \theta_s r_s ds + \int_0^t \theta_s dr_s$$

which is (36) with $\chi = 1$. Minimizing the cost of the contract implies that the θ_s are deterministic. Then, for a constant K, $U = K + \int_0^T \left(\int_s^T e^{-\rho t} dt\right) \theta_s \sigma_s dz_s$, so $\theta_s = Y_s / \left(\int_s^T e^{-\rho t} dt\right)$.

A.5 Microfoundation for the NM Constraint

The analysis assumes that it is optimal for the principal to implement zero manipulation, i.e. $m_t = 0$. Here we offer potential microfoundations for this objective. Assume that the firm's stock yields no dividend except in a period $\tau > L$ after the CEO leaves. The dividend is given by:

$$D_\tau = \exp\left(\sum_{s=1}^{\tau} \eta_s + a_s\right)(1-\mu)$$

where μ depends on the extent of manipulation in way we will soon specify. For simplicity, we normalize η_t such that $E\left[e^{\eta_t}\right] = 1$. At each date $t < \tau$, investors observe the signal:

$$v_t = \sum_{s=1}^{\tau} \eta_s + a_s.$$

Therefore, the rational expectation of firm value is $P_t = e^{-r_f(\tau-t)}E_t\left[D_\tau\right] = e^{-r_f(\tau-t)}e^{v_t}E_t\left[e^{\sum_{s=t+1}^{\tau} a_s^*}\right](1-\mu)$, and the log return is $r_t = \ln P_t/P_{t-1} = \eta_t + a_t - a_t^* + r$.

We now detail the impact of manipulation. We assume that if the probability that the CEO engages in manipulation is greater than zero, then $\mu = \mu_* > 0$, otherwise it is zero. Therefore, we model the cost of manipulation as a fixed cost to firm value: the expectation of even an infinitessimal amount of manipulation lowers firm value by a fixed amount μ_*. This technological assumption gives a tractable way to capture the fact that the possibility of manipulation reduces firm value (e.g. because monitoring is needed to verify accounts or scrutinize investment projects.) Note that the assumption of this cost μ allows us to dispense with the cost $\lambda(m)$ featured in the main paper.

Hence, the loss from allowing manipulation is $\mu_* S$, where S is the firm value at time 0 without manipulation, while the benefit is at most the present value V^{CEO} of the CEO's salary under the scheme avoiding manipulation. Thus, if S is sufficiently large (if it is greater than V^{CEO}/μ_*), it is efficient to design the contract to deter manipulation.

If the CEO engages in manipulation at time t, then firm value rises to $S'_{t+j} = S_{t+j}e^{m_t}$ for $j = 0..., i-1$, and $S'_{t+i} = S_{t+i}$ (manipulation is reversed at $t+i$). This implies that the

returns change from R_t to R_t', with:

$$R_t' = R_t + m_t, \qquad R_{t+i_t} = R_{t+i_t} - m_t, \qquad r_s = R_s \text{ for } s \neq t, t + i_t, \tag{57}$$

For example, $R_t' = \ln\left(S_t e^{m_t}\right) - \ln S_{t-1} = R_t + m_t$, and $R_{t+i_t}' = \ln\left(S_{t+i_t}\right) - \ln\left(S_{t+i_t-1} e^{m_t}\right) = R_{t+i_t} - m_t$. Hence, by the reasoning in the body of the paper, to ensure that the CEO does manipulate, we require $E_t\left[\frac{\partial U}{\partial r_t}\right](m_t) + E_t\left[\frac{\partial U}{\partial r_{t+i}}\right](-m_t) = 0$, i.e. (8).

Another microfoundation is as follows. Instead of assuming a fixed cost μ_*, we assume that a manipulation $m_t > 0$ lowers $\ln D_\tau$ by $\lambda_i^+ m$, and a manipulation $m_t < 0$ lowers it by $-\lambda_i^- m$, where $\lambda_i^\pm$ are some constants. Therefore, manipulation at time t changes returns to::

$$R_t' = R_t + m_t, \qquad R_{t+i_t} = R_{t+i_t} - \left(1 + \varepsilon \lambda_i^\varepsilon\right) m_t, \qquad r_s = R_s \text{ for } s \neq t, t + i_t,$$

where $\varepsilon = sign\,(m)$

The CEO will not engage in manipulation if $E_t\left[\frac{\partial U}{\partial r_t}\right](m) - E_t\left[\frac{\partial U}{\partial r_{t+i}}\right]\left(1 + \varepsilon \lambda_i^\varepsilon\right) m \leq 0$, for small m, which leads to $\left(1 + \lambda_i^+\right) E_t\left[\frac{\partial U}{\partial r_{t+i}}\right] \geq E_t\left[\frac{\partial U}{\partial r_t}\right]$ and $\left(1 - \lambda_i^-\right) E_t\left[\frac{\partial U}{\partial r_{t+i}}\right] \leq E_t\left[\frac{\partial U}{\partial r_t}\right]$. We obtain a series of inequalities which simplify tractably in the case where $1 + \lambda_i^+ = e^{\Lambda i}$ and $\left(1 - \lambda_i^-\right) = e^{-\Lambda i}$. This yields:

$$e^{-\Lambda i} E_t\left[\frac{\partial U}{\partial r_t}\right] \leq E_t\left[\frac{\partial U}{\partial r_{t+i}}\right] \leq e^{\Lambda i} E_t\left[\frac{\partial U}{\partial r_t}\right] \text{ for } 0 \leq i \leq M. \tag{58}$$

They reduce to our NM formulation when $\Lambda \to 0$. We opted for our NM formulation because of its tractability, but the formulation (58) can still be useful in some settings.

References

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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

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