A Calibratable Model of Optimal CEO Incentives in Market Equilibrium[*]

Alex Edmans

Wharton School, University of Pennsylvania

Xavier Gabaix

NYU Stern School of Business and NBER

Augustin Landier

NYU Stern School of Business

Comments welcome
July 24, 2007

Abstract

This paper presents a tractable, neoclassical model of both CEO incentives and total pay in market equilibrium. We embed the traditional principal-agent framework into a competitive assignment model, where managerial talent determines total wages and the size of one's firm. This generates quantitative predictions for the optimal level of CEO incentives and their scaling with firm size. We empirically evaluate the model and show that observed practices are close to our first-best benchmark. In particular, the significant negative relationship between the CEO's effective equity stake and firm size is fully consistent with optimal contracting, and need not reflect rent extraction. While various measures of wealth-performance sensitivity have been used by empiricists, our model proposes that the most appropriate measure is the dollar change in wealth for a percentage change in firm value, scaled by annual pay. Both theory and evidence show that it is independent of firm size, in contrast to alternative measures, and thus comparable between firms and over time.

KEYWORDS: Executive compensation, pay-performance sensitivity, incentives, options, corporate governance, scaling, calibratable corporate finance

JEL CLASSIFICATION: D2, D3, G34, J3

1 Introduction

This paper presents a neoclassical model of the CEO labor market, in which both total salary and its incentive component are simultaneously determined by the market for scarce talent and the magnitude of agency problems. Holding total pay constant, effort considerations determine its division into fixed and performance-sensitive components. To endogenize total pay, we build on previous work (Gabaix and Landier (2008)) by embedding this result into a general equilibrium model of the competitive assignment of CEO talent. The most skilled CEOs are matched with the largest firms and earn the highest salaries, leading to a positive association between total pay and firm size. The absolute level of incentive compensation therefore also varies with size. The model is tractable and calibratable, thus generating quantitative predictions for the optimal level of CEO incentives and their scaling with firm size in a frictionless world.

We explore three main applications of the model. The first is to understand theoretically why wealth-performance sensitivity should *optimally* vary across firms of different size. This issue is important for at least two reasons. It has been widely documented that the CEO's "effective equity stake" (the dollar change in wealth for a dollar change in firm value) is significantly decreasing in firm size (e.g. Jensen and Murphy (1990), Schaefer (1998)). Why is this? One interpretation is that rent extraction is particularly pronounced in large firms, thus allowing incentives to be suboptimally low (e.g. Bebchuk and Fried (2004)). If this argument is correct, the implications are profound. If the CEOs in charge of the largest companies have the weakest incentives to exert effort, then billions of dollars of value may be lost each year. This explanation would also imply a pressing need for intervention: the current system of pay determination is broken, and must be fixed. Our model can be used to evaluate this hypothesis as it provide a quantitative benchmark for how incentives should scale with size under optimal contracting. Unlike other determinants of incentives studied by the literature, size can be measured with little error. This limits our flexibility in calibration, allowing the model to be subject to particularly close empirical scrutiny, and its predictions to be rejectable. We predict that the effective equity stake should have a size elasticity of -2/3, very close to our empirical estimate of -0.58. Therefore, the observed negative relationship between incentives and size need not be evidence of inefficiency – it is exactly what a frictionless model would predict. Similarly, our predicted size elasticity for the dollar-log wealth-performance sensitivity of 1/3 is also empirically supported.

Understanding the scaling of incentive measures with firm size is also important to evaluate the various metrics available to empiricists. We demonstrate both theoretically and empirically that "scaled wealth-performance sensitivity" (the dollar change in wealth for a percentage change in firm value, scaled by annual pay) is invariant to firm size, unlike other commonly used measures. This property may make it particularly attractive for empirical analysis. If the level of incentives is the focus of the empirical study, size independence permits meaningful comparisons across firms or over time. In addition, it ensures that the explanatory power of

the incentives measure does not simply arise because it proxies for size. If other relationships are the focus of the study and incentives are instead used as a control variable, it is desirable to use a "pure" measure of incentives that is not distorted by size.

A second application of the model is to evaluate the absolute level of CEO incentives, again to determine whether they can be consistent with optimal contracting. Jensen and Murphy (1990) find that CEO wealth falls by only $3.25 for every $1,000 loss in shareholder value. This is frequently interpreted as evidence that CEOs are "paid like bureaucrats" and insufficiently punished for failure (see however Hall and Liebman (1998)).

The controversies surrounding incentive pay are centered around magnitudes, not directions. CEO wealth does indeed decline with poor performance; the debate is whether it declines enough. Our model is particularly suited to shed light upon this issue as it generates a quantitative benchmark under optimal contracting. We find that the observed level of wealth-performance sensitivity is not too low if CEO shirking increases his utility by a monetary equivalent no greater than his annual wage. Since it appears plausible that the gains from shirking fall below this upper bound, the level of incentives is also consistent with efficiency.

The intuition behind the model's explanatory power is as follows. The disutility cost of effort is proportional to the manager's consumption and thus his wealth, but its benefit is proportional to firm value. Since firm value is extremely large compared to the manager's wealth, the dollar gains from effort are substantial and so the manager only needs a small equity stake to achieve incentive compatibility. This explains the observed level of CEO incentives. The multiplicative effect of effort also helps to explain the negative relationship between dollar-dollar incentives and size. Since effort has a proportional impact on firm value, the dollar gains from working scale proportionately with size. While the CEO's utility gain from shirking (in dollar terms) also rises with wealth, wages (and thus wealth) only have a $1/3$ elasticity with size. Therefore, dollar-dollar incentives should have a size elasticity of $-2/3$, and so a smaller equity share is sufficient to induce effort among large companies.

Our use of multiplicative functional forms for the costs and benefits of effort was motivated by their particularly attractive properties documented in the macroeconomics literature. This specification, which contrasts with the additive forms typically modeled, turns out to be crucial to the model's explanatory power. While the level of incentives (a single number) can potentially be explained by a number of different models, the requirement to quantitatively explain scalings across firms of different sizes implies a tight constraint on the specifications that can be assumed. This result is potentially applicable to future calibratable models of corporate finance. Our model also departs from traditional frameworks by incorporating an upper bound on the level of CEO effort. This leads to the prediction of a positive relationship between wealth volatility and firm volatility, which we support with new empirical evidence.

We balance the above results by showing that incentive compensation is ineffective at solving agency problems that are additive in firm value, such as perks. Especially for large firms, perk

consumption has such a small percentage effect on firm value that the negative effect on the manager's equity stake is insufficient to induce value maximization. Hence perk consumption has little explanatory power for incentive compensation, and is instead best avoided through active corporate governance. This conclusion is consistent with empirical evidence that corporate governance does affect firm value, over and above its effect on the CEO compensation contract.

A quite separate third application is to analyze further the determinants of CEO pay. In Gabaix and Landier (2008), some of us examined the impact of scarcity of talent and firm size. Here we investigate the additional importance of disutility of effort and risk. Cross-sectional differences in these parameters naturally lead to between-firm variation in wages. However, we show that market-wide changes in these variables have negligible impact on the pay of the most talented CEOs. Since the pay of top CEOs is principally driven by firm size and the scarcity of CEO talent, it is little affected by compensation for effort or risk. Hence effort and risk explain pay differences along the cross section, but do not affect CEO pay in the aggregate.

By endogenizing both total pay and incentives together, our general equilibrium approach generates results not achievable by simply combining the conclusions of separate models of pay and incentives. In particular, it allows us to understand the factors that do *not* determine CEO pay. For example, we show that the CEO's incentives can be determined independently of the level of his overall compensation – the latter is entirely driven by forces in the managerial labor market. Therefore, high overall pay does not come from the requirement to give the CEO strong incentives, but rather from the marginal productivity of CEO talent in market equilibrium. Incentive considerations change the sensitivity of pay to performance, but not the expected pay. More generally, a single model endogenizing both total pay and incentives may be useful for further work in executive compensation. For empiricists, it constitutes a benchmark against which to quantify inefficiencies in either dimension of observed compensation. For theorists, it is a simple equilibrium framework upon which future, more complex models can potentially be built.

This paper builds on the empirical literature quantifying CEO incentives, and in particular their relationship with firm size. Jensen and Murphy's (1990) seminal study showed that CEOs' dollar-dollar wealth-performance sensitivity is economically very small, particularly for large firms. Schaefer (1998) later confirmed this negative scaling. Hall and Liebman's (1998) more recent evidence illustrates that the recent rise in stock option compensation has significantly increased incentives since the Jensen and Murphy sample period. However, in the absence of an efficient benchmark, we cannot evaluate whether they are now "high enough."

The most closely related theory papers are calibrations of the CEO incentive problem. While the focus of our calibrations is the scaling of CEO incentives with size, Dittmann and Maug (2007) and Armstrong, Larcker and Su (2007) explore the optimal structure of compensation, in particular whether options are a feature of an efficient remuneration package. Garicano and

Hubbard (2005) also calibrate a high-talent labor market, the market for lawyers. Gayle and Miller (2007) explore the contribution of moral hazard to the rise in CEO pay. Baker and Hall's (2004) calibrations estimate the relationship between CEO productivity and firm size. They are the first to recognize that this relationship affects the relevant measure of wealth-performance sensitivity for use in empirical analysis. An analysis of percentage equity holdings implicitly assumes the effect of a CEO's actions is constant in dollar terms, but if the CEO's impact is linear in firm size, the relevant variable is the manager's dollar stake. However, neither measure is stable across size, unlike our proposed metric. Their purpose is to estimate the scaling of managerial productivity with size, not the effect of size on incentives or the optimality of existing practices.

Our paper differs from the above papers owing to its contrasting objectives (the effect of size on incentives) and its modeling approach (general equilibrium with multiplicative functional forms). The general equilibrium framework also differentiates our paper from Haubrich (1994), who identifies the parameter values in the traditional principal-agent model that would be consistent with the 0.325% effective equity stake found by Jensen and Murphy (1990). He notes that the large number of free variables makes it relatively easy to match one moment. We evaluate the ability of a simple neoclassical model to explain the level of incentives, and their scaling with firm size and firm volatility.

In contemporaneous work, Baranchuk, Macdonald and Yang (2007) and Falata and Kadyrzhanova (2007) also model the equilibrium determination of both total pay and its incentive component. The former study focuses on the effect of product market conditions on CEO compensation; the latter analyzes the effect of industry dynamics (in particular the importance of industry structure and a firm's position versus its industry peers.)

A separate literature to which this paper relates examines the optimality of CEO compensation practices. Bebchuk and Fried (2004) argue that certain features of CEO pay reflect rent extraction; see Kuhnen and Zwiebel (2007) for a recent model of hidden pay. However, others have argued that such features may in fact be efficient. Examples include the level of total pay (Gabaix and Landier (2008)), severance pay (Almazan and Suarez (2003), Manso (2006), Inderst and Mueller (2006)), pensions (Edmans (2007)), and perks (Rajan and Wulf (2006)).

This paper is organized as follows. In Section 2 we present a parsimonious model of wages and incentives in general equilibrium where the CEO is risk-neutral. In Section 3 we present empirical evidence that quantitatively supports the model's main predictions, in particular the stability of the scaled wealth-performance sensitivity across firm size. Section 4 studies the optimal contract for a risk-averse CEO. Section 5 illustrates the necessity of certain features of our model to generate empirically consistent predictions, as well as considering further extensions, and Section 6 concludes.

2   The Basic Model

Section 2.1 derives the optimal division of CEO compensation into stock and cash salary, in a partial equilibrium analysis that takes total compensation as given. Section 2.2 embeds this analysis into a general equilibrium where total pay is endogenously determined, and the implications for pay-performance sensitivity are presented in Section 2.3. Section 2.4 illustrates that these results naturally extend to measures of wealth-performance sensitivity, where CEO incentives are principally provided by existing security holdings, rather than flow compensation. Since our objective is to provide calibratable predictions, all of the above results are derived with a deliberately parsimonious model where the CEO is risk-neutral, the effort decision is binary, and the contract is restricted to comprise cash and shares. A second reason for starting with risk neutrality is that it gives us one fewer degree of freedom in calibration. Since risk aversion is difficult to measure accurately, a wide range of inputs can be used, thus making it easier to explain the data. We wish to see the extent to which a neoclassical model can match the data without assuming risk aversion. Section 4 will later show that our predictions are robust to relaxing these assumptions.

2.1   Incentive Pay in Partial Equilibrium

The CEO's objective function is:

$$U = E\left[c \cdot g\left(e\right)\right], \tag{1}$$

where c is the CEO's monetary compensation and $e \in \{-1, 0\}$ denotes CEO effort. We normalize $g\left(0\right) = 1$ and set $g\left(-1\right) = 1/\left(1-\Lambda\right)$, where $\Lambda \in [0, 1)$ parameterizes the disutility of effort. The CEO is subject to limited liability ($c \geq 0$) and has a reservation utility of w, the wage available in alternative employment. This is endogenized in Section 2.2.

Equation (1) is generalizable to other multiplicative forms, such as $E\left[\left(cg\left(e\right)\right)^{\alpha}\right]$. In macroeconomics, multiplicative functional forms (such as $u = \left(cg\left(\text{hours worked}\right)\right)^{\alpha}$) calibrate particularly well across different levels of wealth.[1] This motivates our specification choice here. In Section 5.1 we show that the additive functional forms more commonly used in corporate finance, such as $E\left[c^{\alpha}\right] - g\left(e\right)$, are less suited for calibration.

The initial stock price is P, and the end-of-period stock price is given by

$$P_1 = P\left(1 + Le + \eta\right), \tag{2}$$

where η is stochastic noise with mean 0. Low effort ($e = -1$) reduces firm value by a fraction

---

[1]For example, consider the labor supply l of a worker living for one period, with a wage w, consumption $c = wl$, and utility $v\left(c, l\right)$. He solves $\max_l v\left(wl, l\right)$. If utility is $v\left(c, l\right) = \phi\left(cg\left(l\right)\right)$, then the problem is $\max_l \phi\left(wl\ g\left(l\right)\right)$, and the optimal labor supply l is independent of w. This is a desirable property so that the model does not predict labor supply having diverging trends over time.

L. We assume that $SL > w\Lambda$, where S is the firm's market capitalization[2]: the firm value gains from high effort exceed the manager's disutility, and so it is optimal to elicit effort.[3]

This paper defines effort broadly, to apply to any decision that increases firm value but involves a non-pecuniary cost to the manager. In the literal interpretation, $e = 0$ represents "working" and $e = -1$ is "shirking". A second interpretation is project or strategy choice, where $e = 0$ is the first best project and $e = -1$ yields the CEO "private benefits" (e.g. Aghion and Bolton (1992)). The effects of effort or project choice are plausibly multiplicative in firm value, explaining the formulation in equation (2). However, the effect of "perks", such as a corporate jet or "vanity expenditures" such as sponsorship of a sporting event, is fixed in dollar terms and thus additive to firm value. We consider such actions in Section 5.3.

The CEO's compensation c is composed of a fixed cash salary $f \geq 0$, and ν shares:[4]

$$c = f + \nu P_1. \tag{3}$$

The optimal contract elicits high effort ($e = 0$) and pays the CEO his reservation wage, i.e. $E[c] = w$. Since the manager is risk neutral (for $c > 0$), many compensation packages are optimal. In Proposition 1 below, we derive the contract that minimizes the number of shares given to the manager, since this would be optimal if the CEO had vanishingly small but positive risk aversion.

Proposition 1 *(CEO incentive pay in partial equilibrium). Fix the manager's expected pay at w and assume $L > \Lambda$ (the cost of effort is not too strong). The optimal contract comprises a fixed base salary, f^*, and $\nu^* P$ worth of shares, with:*

$$\nu^* P = w \frac{\Lambda}{L}, \tag{4}$$

$$f^* = w \left(1 - \frac{\Lambda}{L}\right), \tag{5}$$

where L is the percentage decrease in firm value if the manager shirks, and Λ is the manager's disutility of effort. The manager's realized compensation is:

$$c = w \left(1 + \frac{\Lambda}{L}(r - E[r])\right), \tag{6}$$

---

[2]For simplicity, we assume an all-equity firm. If the firm is levered, S represents the aggregate value of the assets of the firm (debt plus equity) and P denotes the aggregate value per share.

[3]The proof is as follows. If the manager works, he is paid w and firm value (net of wages) is $S - w$, leading to total surplus of S. If the manager shirks, he is paid $w(1 - \Lambda)$ (to keep his utility at w). Firm value (net of wages) is $S(1 - L) - w(1 - \Lambda)$ and total surplus is $S(1 - L) + w\Lambda$. Hence total surplus is higher if the manager works if and only if $SL > w\Lambda$.

[4]Section 4 extends the model to general contracts under risk aversion. In the online appendix (Appendix D) we show the results are unchanged by generalizing to other instruments, such as options, while retaining risk neutrality.

where $r = P_1/P - 1$ is the firm's stock market return.

In the optimal contract described by Proposition 1, realized CEO compensation is not indexed to the market and CEOs are rewarded for luck. Therefore, the empirical observation of these practices (e.g. Bertrand and Mullainathan (2001)) need not be inconsistent with optimal compensation. This result stems from the assumption that the CEO is risk neutral and so the informativeness principle of Holmstrom (1979) does not apply. In reality, CEOs likely exhibit some degree of risk aversion, providing a motive for indexation. This is counterbalanced by the costs of additional complexity in writing indexed contracts. Reality likely reflects a trade-off between these two factors.

2.2   Incentive Pay in Market Equilibrium

We now embed the previous analysis into a market equilibrium where the equilibrium wage w is endogenously determined. We directly import the model of Gabaix and Landier (2008) ("GL"), the essentials of which we review in the Appendix. There is a continuum of firms of different size and managers with different talent. Since talented CEOs are more valuable in larger firms, the nth most talented manager is matched with the nth largest firm in competitive equilibrium, and earns the following competitive equilibrium pay:

$$w(n) = D(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma - \beta/\alpha}, \tag{7}$$

where $S(n)$ is the size of firm n, n_* is the index of a reference firm (e.g. the median firm in the economy), $S(n_*)$ is the size of that reference firm, and $D(n_*)$ is a constant independent of firm size. In particular, CEOs at large firms earn more as they are the most talented, with a pay-firm size elasticity of $\rho = \gamma - \beta/\alpha$ that GL calibrate to 1/3.

GL only specify the total compensation that the CEO must be paid in market equilibrium. We now seamlessly incorporate the incentive results of Section 2.1 to determine the form of compensation. We allow L and Λ to differ across firms, and so index them L_n and Λ_n. We do not need to make any assumptions on how L_n or Λ_n vary with n: as long as $L_n > \Lambda_n$ for each firm, effort can be induced by the incentive contract. Since there is no shirking, the "baseline" firm value remains at S, as in GL. The equilibrium incentive pay is analogous to Proposition 1:

Proposition 2 *(CEO incentive pay in market equilibrium). Assume $\forall\, n, L_n > \Lambda_n$ (the cost of effort is not too strong). Let n_* denote the index of a reference firm. In equilibrium, the manager of index n runs a firm of size $S(n)$, and is paid an expected wage:*

$$w(n) = D(n_*) S(n_*)^{\beta/\alpha} S(n)^{\gamma - \beta/\alpha}, \tag{8}$$

where $S(n_*)$ is the size of the reference firm and $D(n_*) = -n_* T'(n_*) / (\alpha\gamma - \beta)$ is a constant independent of firm size. The optimal contract pays manager n a fixed base salary, f_n^*, and

$\nu_n^* P_n$ worth of shares, with:

$$\nu_n^* P_n = w(n) \frac{\Lambda_n}{L_n},$$

$$f_n^* = w(n) \left(1 - \frac{\Lambda_n}{L_n}\right),$$

where L_n is the percentage decrease in firm value if the manager shirks, and Λ_n is the manager's disutility of effort. The manager's realized compensation is:

$$c(n) = w(n) \left(1 + \frac{\Lambda_n}{L_n}(r(n) - E[r(n)])\right),$$

where $r(n) = P_{1n}/P_n - 1$ is the firm's stock market return during the period.

To our knowledge, the above Proposition yields the first closed-form solution for a market equilibrium determination of optimal CEO incentives, in a model where CEOs have different talents. The most similar antecedent is Himmelberg and Hubbard (2000), which does not have closed forms.

Note that the total level of pay $w(n)$ is determined entirely by the CEO's marginal product, and is independent of incentive considerations. The latter only affects the division of total pay into cash and stock components. Hence high pay is not "justified" by the need to reward CEOs for good performance, or to compensate them for the risk associated with incentive compensation: CEOs are risk-neutral in our model. As in GL, high levels of pay are entirely justified by scarcity in the market for talent, not by incentive considerations. Simply put, total compensation is driven by "pay-for-talent", not "pay-for-performance". Empirically observing high pay despite poor firm performance need not automatically imply inefficiency, since in a competitive market, high pay may have been necessary to attract a skilled manager.[5] As long as pay would have been even higher had the manager delivered stronger performance, it can be consistent with optimal contracting.

2.3 Pay-Performance Sensitivities in Market Equilibrium

The empirical literature uses a variety of measures for pay-performance sensitivity. These are defined below (we suppress the dependence on firm n for brevity).

Definition 1 *Let c denote realized compensation, w the expected pay, S the market value of*

---

[5] For example, the large severance package given to Robert Nardelli of Home Depot appears ex post inefficient, but it may have been necessary ex ante to attract a manager of his talent.

the firm, and r the firm's return. We define the following pay-performance sensitivities:

$$b^I = \frac{\partial c}{\partial r}\frac{1}{w} = \frac{\Delta \ln Compensation}{\Delta \ln Firm\ Value} \tag{9}$$

$$b^{II} = \frac{\partial c}{\partial r}\frac{1}{S} = \frac{\Delta\$Compensation}{\Delta\$Firm\ Value} \tag{10}$$

$$b^{III} = \frac{\partial c}{\partial r} = \frac{\Delta\$Compensation}{\Delta \ln Firm\ Value}. \tag{11}$$

b^I is used (or advocated) by Murphy (1985) and Rosen (1992); b^{II} by Demsetz and Lehn (1985), Yermack (1995) and Schaefer (1998); and b^{III} by Holmstrom (1992). The next Proposition derives predictions for these quantities, in the case where $\Lambda_n = \Lambda$ and $L_n = L$ across all firms.[6]

Proposition 3 *(Pay-performance sensitivities). Equilibrium pay-performance sensitivities are given by:*

$$b^I = \frac{\Lambda}{L} \tag{12}$$

$$b^{II} = \frac{\Lambda}{L}\frac{w}{S} \tag{13}$$

$$b^{III} = \frac{\Lambda}{L}w, \tag{14}$$

where w is given by (7).

Share-based compensation can be implemented in a number of forms, such as stock grants, bonuses and reputational concerns. If the incentive component is implemented purely using shares, these sensitivities have natural interpretations. b^I represents the dollar value of the CEO's shares as a proportion of the CEO's total pay, b^{II} is the percentage of shares outstanding held by the CEO, and b^{III} represents the dollar value of the CEO's shares. If the incentive component is implemented using other methods, the above coefficients constitute the "effective" share ownership.

Proposition 4 *(Scaling of pay-performance sensitivities with firm size). Let ρ denote the cross-sectional elasticity of expected pay to firm size: $w \propto S^\rho$. For instance, in GL, $\rho = \gamma - \beta/\alpha$. The pay-performance sensitivities scale in the following way:*

1. In the cross-section, b^I is independent of firm size:

$$b^I \propto S^0.$$

---

[6]We make this assumption to maintain the simplicity of our model and limit our degrees of freedom in calibration. The model can be extended to allow the effort parameters to vary across firms, as in Baker and Hall (2004).

2. In the cross-section, b^{II} scales as $S^{\rho-1}$:

$$b^{II} \propto S^{\rho-1}.$$

3. In the cross-section, b^{III} scales as S^{ρ}:

$$b^{III} \propto S^{\rho}.$$

In particular, in the calibration $\rho = 1/3$ used in GL,

$$b^{I} \propto S^{0}, \; b^{II} \propto S^{-2/3}, \; and \; b^{III} \propto S^{1/3}. \tag{15}$$

Proposition 5 *(Dependence of pay-performance sensitivities on the size of the reference firm).*
Let n_* denote the index of a reference firm and $S(n_*)$ its size. The pay-performance sensitivities
scale with $S(n_*)$ in the following way:

$$\begin{aligned} b^{I} &\propto S^{0} S(n_*)^{0} \\ b^{II} &\propto S^{-(1-\rho)} S(n_*)^{\gamma-\rho} \\ b^{III} &\propto S^{\rho} S(n_*)^{\gamma-\rho} . \end{aligned}$$

where γ is the elasticity of CEO impact in GL (equation (40)). In particular, in the calibration
$\rho = 1/3, \gamma = 1$, used in GL,

$$b^{I} \propto S^{0} S(n_*)^{0} , \; b^{II} \propto S^{-2/3} S(n_*)^{2/3} , \; and \; b^{III} \propto S^{1/3} S(n_*)^{2/3} .$$

Table 1 summarizes our results for the different measures of pay-performance sensitivity.

Insert Table 1 about here

Propositions 4 and 5 imply that the log-log measure of pay-performance sensitivity is in-
dependent of both firm size and the size of reference firms. The intuition is as follows. In
our model, effort has a percentage effect on both firm value and the CEO's utility. Since this
percentage is constant across firms, the required %-% (or log-log) incentives to achieve incentive
compatibility should be constant across size.

This result suggests that b^{I} is the most appropriate measure of CEO incentives to use when
comparing between firms or different time periods. Note that this proposal stems from our
assumption that effort has multiplicative costs and benefits. Baker and Hall (2004) show that,
under different assumptions, b^{II} or b^{III} may be appropriate. Which assumptions are closest to

11

reality is therefore an empirical question. Section 3 presents evidence that supports the model's prediction that b^I is stable and that other measures are size-dependent.

Proposition 4 also predicts that b^{II} should decline with firm size, a relationship widely documented empirically. Since $b^{II} = b^I \frac{w}{S}$ and the wage w in market equilibrium only scales with $S^{1/3}$, b^{II} is predicted to scale with $S^{-2/3}$. Existing interpretations of this stylized fact are greater managerial entrenchment and inefficiency in large firms (Bebchuk and Fried (2004)), stronger political constraints on high pay in large, visible firms (Jensen and Murphy (1990)), greater volatility thus imposing higher risk on the CEO (Schaefer (1998)), and wealth constraints limiting the percentage of a large firm that a CEO can hold (Demsetz and Lehn (1985)). Our explanation does not rely on any of these constraints; b^{II} optimally falls with size because managerial effort is multiplicative in firm value and thus substantially increases the dollar value of a large firm. Therefore, a smaller percentage equity holding is required to induce effort: applied to a large dollar value change, this creates a sufficient incentive to work. It is efficient for CEOs of large firms to be "paid like bureaucrats", as found by Jensen and Murphy (1990). This point has been previously noted by Hall and Liebman (1998) and modeled by Baker and Hall (2004); we form a quantitative prediction for this scaling in market equilibrium.

Finally, b^{III} is the effective dollar equity stake. Section 2.1 shows that this should be proportional to total pay. However, since total pay is less than proportional to firm size (it scales with $S^{1/3}$), dollar equity holdings should also be less than proportional to firm size.

2.4 Wealth-Performance Sensitivities in Market Equilibrium

Thus far, we have assumed the CEO's incentives stem purely from his flow compensation. However, for many CEOs, the vast majority of incentives stem from changes in the value of existing holdings of stock and options (see Hall and Liebman (1998), Core, Guay and Verrecchia (2003) among others). Appendix C.1 presents a full model that extends the previous results to a multiperiod setting. The key results are summarized here.

Replacing flow compensation in the numerator of Definition 1 with the overall change in wealth yields the following definitions of *wealth*-performance sensitivity:

Definition 2 *Let W denote total CEO wealth (including NPV of future consumption), w the expected flow pay, S the market value of the firm, and r the firm's return. We suppress the dependence on firm n for brevity and define the following wealth-performance sensitivities:*

$$B^I = \frac{\partial W_t}{\partial r_t} \frac{1}{w_t} = \frac{\Delta \$Wealth}{\Delta \ln Firm\ Value} \frac{1}{\$Wage} \tag{16}$$

$$B^{II} = \frac{\partial W_t}{\partial r_t} \frac{1}{S_t} = \frac{\Delta \$Wealth}{\Delta \$Firm\ Value} \tag{17}$$

$$B^{III} = \frac{\partial W_t}{\partial r_t} = \frac{\Delta \$Wealth}{\Delta \ln Firm\ Value}. \tag{18}$$

B^{II} is used by Jensen and Murphy (1990). Hall and Liebman (1998) report both B^{II} and B^{III}, as well as a variant of B^I where the denominator is flow compensation w_t *plus* the median return applied to the CEO's existing portfolio of shares and options.[7]

Multiplying the pay-performance sensitivities in Proposition 5 by $\frac{W}{w}$ gives the following magnitudes for wealth-performance sensitivities:

Proposition 6 *(Wealth-performance sensitivities). Let W denote total CEO wealth (including NPV of future consumption) and w the expected flow pay. Then:*

$$B^I \quad = \quad \frac{\Lambda}{L}\frac{W_t}{w_t} \tag{19}$$

$$B^{II} \quad = \quad \frac{\Lambda}{L}\frac{W_t}{S_t} \tag{20}$$

$$B^{III} \quad = \quad \frac{\Lambda}{L}W_t. \tag{21}$$

The scalings with firm size S and the size of the reference firm S_* are as in Propositions 4 and 5.

Proposition 6 predicts that all three measures of wealth-performance sensitivity are higher for wealthier CEOs. This has been empirically confirmed by Becker (2006) for B^{II} and B^{III} (he does not investigate B^I). Becker's explanation is that risk aversion declines with wealth, therefore rendering incentive pay less costly. Our model offers a different explanation that does not rely on risk aversion. The multiplicative utility function means that shirking and consuming are complementary goods, which is realistic since free time is required to enjoy consumption. Higher wealth raises current consumption and thus the utility gains from shirking. Pay-performance sensitivity must therefore rise to continue to induce effort.

3 Empirical Evaluation

This section calculates empirical measures of wealth-performance sensitivity and assesses the extent to which current practices are consistent with our neoclassical benchmark. Section 3.1 shows that the data is quantitatively consistent with the model's predictions for the scalings of incentives with firm size. In particular, B^I is independent of size and we therefore propose it as the preferred empirical measure of incentives. Section 3.2 calibrates the level of incentives and show that they can be fully consistent with efficiency.

---

[7]Note that we scale B^I by the wage, not by wealth which may seem more intuitive. The reason is data limitations: in the U.S., the only wealth data we have is on the CEO's security holdings in his own firm. Therefore, measured wealth will mechanically have a (close to) constant firm value elasticity – for example, if he holds stock and no options, $\frac{\partial W_t}{\partial r_t}\frac{1}{W_t}$ would equal 1.

3.1 CEO Incentives and Firm Size

Proposition 5 summarized the model's predictions for the cross-sectional scaling of incentive pay with firm size. Our model predicts that the dollar-dollar wealth-performance sensitivity, B^{II}, should optimally decline with firm size. This directional association has been consistently documented by a number of existing studies, such as Demsetz and Lehn (1985), Jensen and Murphy (1990), Gibbons and Murphy (1992), Schaefer (1998), Hall and Liebman (1998) and Baker and Hall (2004). Moreover, our calibratable framework allows us to derive quantitative predictions of the elasticity of b^{II} with respect to size. Specifically, $\gamma - \beta/\alpha = 1/3$ (as found by GL) implies an elasticity of $-2/3$. Consistent with our model, Schaefer finds $B^{II} \sim S^{-\xi}$, with $\xi \simeq 0.68$.[8] Existing research is also consistent with the model's prediction that B^I is independent of size (Gibbons and Murphy (1992), Murphy (1999)). We do not know of any studies that investigate the link between B^{III} and size.

However, prior findings cannot be interpreted as conclusive support of the model. Some of the above studies focus on the compensation *flows* (salary, bonus and new grants of stock and options) but do not have full data on the CEO's *stock* of shares and options which provide the vast majority of CEO incentives.

We therefore conduct our own empirical tests of the model, using measures of wealth-performance sensitivity. We merge Compustat with ExecuComp (1992-2005) and select the largest 500 firms in aggregate value (debt plus equity) in each year.[9] We calculate the wealth-performance sensitivities as follows:

$$B^I \quad = \quad \frac{1}{w_t}\left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \tag{22}$$

$$B^{II} \quad = \quad \frac{1}{S_t}\left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \tag{23}$$

$$B^{III} \quad = \quad \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \tag{24}$$

We use the Core and Guay (2002a) methodology to estimate the option deltas. (Appendix A describes our calculations in further detail.) Controlling for year and industry fixed effects, and clustering standard errors at the firm level, we estimate the following elasticities:

$$\ln(B^I_{i,t}) \quad = \quad \alpha + \beta \times \ln(S_{i,t})$$
$$\ln(B^{II}_{i,t}) \quad = \quad \alpha + \beta \times \ln(S_{i,t})$$
$$\ln(B^{III}_{i,t}) \quad = \quad \alpha + \beta \times \ln(S_{i,t}).$$

---

[8]This ξ is taken from Table 4 of Schaefer (1998), and is equal to $1 - 2(\phi - \gamma)$ using his notation. We average over his four estimates of ξ. Note that Schaefer estimates a non-linear model that is closely related to ours, but not identical, so his findings only constitute weak support.

[9]Our results are very similar if we use sales as a measure of firm size, and if we select the top 1000 or 200 firms.

Table 2 illustrates the results, which are consistent with the predictions of equation (15). Specifically, B^I is independent of firm size: the coefficient of 0.06 is slightly less than its standard deviation. B^{II} (B^{III}) have size elasticities of -0.58 (0.42), statistically indistinguishable from the model's prediction of $-2/3$ (1/3). Our model can therefore *quantitatively* explain the size elasticities of all three measures of wealth-performance sensitivity.

In unreported results, adding the Gompers, Ishii and Metrick (2003) governance index as an explanatory variable yields a coefficient of -0.057, statistically significant at just greater than the 1% level. The standard deviation of the governance index is 2.7, implying that a one standard deviation rise in the index (i.e. a worsening of governance) is associated with B^I falling by 15%.

<div align="center">

Insert Table 2 about here

</div>

The empirical literature has used a wide variety of measures of CEO incentives, but there has been limited theoretical guidance over which measure is appropriate. A notable exception is Baker and Hall (2004), who show that the optimal measure depends on the scaling of CEO productivity with firm size. If productivity is constant in dollar terms regardless of firm size, b^{II} (or B^{II}) is appropriate as it is size-invariant; if it is linear in firm size, b^{III} (or B^{III}) is the correct measure as it becomes size-invariant. However, their calibrations estimate the size-elasticity of CEO productivity of 0.4, in between the two extremes, suggesting that both measures may be problematic.

Our model predicts that B^I is independent of firm size. While this stemmed from our assumption that effort has multiplicative costs and benefits, Table 2 empirically confirms its size invariance (thus supporting our modeling assumptions) as well as the size dependence of B^{II} and B^{III}. This property may render B^I an attractive measure of CEO incentives in a number of empirical applications. If the level of incentives is the focus of the empirical study, size independence permits meaningful comparisons across firms or over time. In addition, it ensures that the explanatory power of the incentives measure does not simply arise because it proxies for size. If other relationships are the focus of the study and incentives are instead used as a control variable, it is desirable to use a "pure" measure of incentives undistorted by size.

3.2 The Level of CEO Incentives

We now use our model to assess whether currently observed levels of wealth-performance sensitivity are consistent with efficiency. Our primary measure is the log-log pay-for-performance sensitivity; the other measures are mechanical transformations. The model predicts $B^I = \frac{\Lambda}{L}\frac{W}{w}$ (equation (49)). The median B^I for 2003-5 is 11 and it is stable over this period.[10]

---

[10]Hall and Liebman (1998, Table VIII) estimate $B^I = 3.9$. Their denominator includes not only flow compensation but also the expected appreciation of the CEO's stock and options.

Unfortunately, L and particularly Λ are difficult to measure precisely. Our aim is to identify reasonable combinations of Λ and L implied by $B^I \simeq 11$. By seeing whether these values appear plausible, we can assess whether the empirically observed pay-for-performance sensitivity is necessarily suboptimal. An analogy is the equity premium puzzle of Mehra and Prescott (1985). Risk aversion is difficult to measure precisely, so the authors posit an "admissible region" of plausible values. Their calibrated level of risk aversion falls outside this region, giving rise to the puzzle.

L is the percentage amount by which firm value decreases if the CEO shirks. A natural starting point is the average takeover premium of 30%. However, the takeover premium can be motivated by factors other than managerial misbehavior, such as synergies or undervaluation. Since a high input for L makes it easier to match the B^I found in the data, we conservatively set $L \simeq 10\%$. We therefore calibrate

$$\Lambda = \frac{B^I \cdot L}{\frac{W}{w}} = \frac{11 \cdot 0.1}{\frac{W}{w}} = 1.1\frac{w}{W}.$$

Shirking increases the CEO's utility by a fraction $\Lambda = 1.1\frac{w}{W}$ of his wealth, i.e. \$1.1w in dollar terms. Rounding down to be conservative, the "private benefits of shirking" can increase the CEO's utility by an amount no greater than his annual salary.

To turn this into numerical amounts, the median pay of the 500 CEOs in our sample averaged \$5.7 million for 2003-5. The utility from shirking can therefore be no higher than \$5.7 million. Since this is a high upper bound, it is likely that the actual utility from shirking falls within the "admissible region" and so we cannot conclude that current practices are inefficient. Note that the above calibration does not require an estimation of W/w, since it cancels out. The only degree of freedom we have in our calibration is the input L.

To calibrate Λ as a percentage, we would need to estimate W/w. Unfortunately, there is no data available on the wealth W of CEOs.[11] However, ExecuComp provides data on a CEO's financial wealth in his own firm. For 2003-5, we estimate a median value of (Financial wealth in the firm) / (Pay) equal to 12. We assume that the CEO's wealth in his own firm is half his total financial wealth, and that his human wealth (NPV of future wages) approximately equals his entire financial wealth. This leads to an estimate of W/w of 48. We therefore have

$$\Lambda = 1.1\frac{w}{W} \simeq \frac{1.1}{48} = 0.023.$$

This means that, if the CEO shirks, his utility increases by an amount equivalent to 2.3% of his wealth. Section 4 uses the quantity Λ/L, which we calibrate to be approximately 1/4.

Since B^{II} and B^{III} are mathematically linked to B^I, our ability to explain B^I means that the model can also match the measures of wealth-performance sensitivity more commonly used

---

[11] We thank David Yermack for discussions on this point.

by empiricists. For example, $B^{II} = B^I \frac{w}{S}$. The median size of the top 500 firms, averaged across 2003-5, is \$15.2 billion. $B^I = 11$ is therefore consistent with a Jensen-Murphy semi-elasticity of $B^{II} = 11 \times (\$5.7 \text{ million}) / (\$15.2 \text{ billion})$. This represents a wealth rise of \$4.12 for a \$1,000 increase in firm value, close to our figure of \$3.68.[12]

4 Extended Model with Risk Aversion and Optimal Contracts

We have thus far assumed a risk-neutral CEO, a binary effort decision, and limited our instruments to cash and shares. This was to maximize the model's tractability and thus calibratability. This section introduces risk aversion and multiple effort levels into a continuous time setup, and derives the optimal contract without restricting the contracting space. In addition to testing the robustness of our predictions, the extended model also allows us to analyze the effect of risk on compensation. Section 4.1 considers the extended model in partial equilibrium, and in Section 4.2 we embed it in market equilibrium.

4.1 Partial Equilibrium

We use a continuous time framework because, as known since Holmstrom and Milgrom (1987), this leads to contracts that are simpler and more robust than those that come from a discrete time analysis.

We consider a period of length T. At each date t within this period, the CEO exerts effort e_t, where $e_t \in [\underline{e}, \overline{e}]$. The stock price evolves according to: $dP_t/P_t = (r_f + \pi + L(e_t - \overline{e})) dt + \sigma dz_t$. The CEO's utility function is given by:

$$U = E\left[u\left(c_T \exp\left(-\Lambda \int_0^T e_t dt\right)\right)\right], \tag{25}$$

with c_T is the consumption at T, $u(c) = c^{1-\Gamma}/(1-\Gamma)$ for $\Gamma \geq 0$, $\Gamma \neq 1$, $u(c) = \ln c$ for $\Gamma = 1$.

The above utility function (25) preserves and generalizes (1) in a number of ways. First, $\Gamma > 0$ measures the CEO's relative risk aversion. Second, effort and consumption continue to affect each other multiplicatively rather than additively. Third, we incorporate multiple periods and allow the CEO to choose a different effort e_t in each period, where e_t depends on the information available up to time t (i.e. is an adapted process). In addition, e_t is no longer a binary variable.

We assume that the maximum level of effort, $e_t = \overline{e}$, maximizes total surplus. As before, this is optimal because the firm (and thus the benefit from effort) is very large compared to the

---

[12]This figure is smaller than the \$5.29 reported by Hall and Liebman because we are considering only the top 500 firms. Across the whole sample, the average median for 2003-5 is \$8.92.

CEO (and thus the cost of effort). The cost of effort now comprises both the direct disutility as well as the inefficient risk sharing that results from incentivizing the manager to exert effort.

The CEO has a reservation utility $u(\underline{w})$ given by the competitive market, and we seek the optimal (unrestricted) contract that implements $e_t = \bar{e} \; \forall \; t$, solves the participation constraint $U \geq u(\underline{w})$, and has the minimum cost $E[c_T]$ to the firm.[13] The solution is derived in Appendix B and stated below.

Proposition 7 *(Optimal contract in the extended model, partial equilibrium). Let $u(\underline{w})$ denote the CEO's reservation utility. The optimal unrestricted contract is as follows. At $t = 0$ the CEO is given wealth $W_0 = \underline{w}a_T$, where a_T is given by (43) . It is invested in a continuously rebalanced account, where a fraction $\theta = \Lambda/L$ is invested in the firm at all times, and the remainder is in the riskless asset. The CEO's terminal wealth is:*

$$\ln \frac{W_T}{W_0} = \theta \ln \frac{P_T}{P_0} + b_T. \tag{26}$$

where b_T is given in (44).

Furthermore, for other $\theta > \Lambda/L$, any contract of the above type is incentive compatible and satisfies the CEO's participation constraint, but it costlier to the firm.

The link with the optimal contract in Section 2 is as follows. In discrete time, changes in log CEO wealth must be proportional to changes in log firm value, with a constant of proportionality of at least $\theta = \Lambda/L$. In continuous time, the same applies at every instant. Hence, at the end of the period, the log change in CEO wealth is proportional to the log change in firm value, with a sensitivity θ.[14] Equation (26) means that final compensation is proportional to the stock price to the power θ. (In the special case where $\Lambda = L$, $\theta = 1$ and the CEO is compensated entirely in stock.)

4.2 Market Equilibrium

4.2.1 Firms Identical Except For Size

We now work out the market equilibrium with risk averse CEOs, using the optimal contract of the previous section. The CEO's terminal utility is:

$$U = u\left(W_0 e^{\left(r_f + \theta\pi - \Lambda\bar{e} - \Gamma\theta^2\sigma^2/2\right)T}\right) \tag{27}$$

For simplicity, we now take the interest rate r_f and risk premia to be 0, i.e., $r_f = \pi = 0$, and the period length to be $T = 1$. Let w denote W_0 for ease of notation. The CEO's utility

---

[13]More precisely, the firm minimizes the market value of the compensation, i.e. $E^Q c_T$, where Q is the risk-neutral probability. This leads to the same solution.

[14]The terms proportional to T in equation (26) simply reflect an adjustment for time value and risk aversion, and are of little interest.

U from (27) is:

$$U = u\left(w\exp\left(-\chi\right)\right) \tag{28}$$

where

$$\chi = \Lambda\bar{e} + \frac{\Gamma\Lambda^2\sigma^2}{2L^2} \tag{29}$$

denotes the "equivalent variation", i.e. the utility loss suffered by the manager by exerting effort (the $\Lambda\bar{e}$ term) and bearing risk (the $\frac{\Gamma\Lambda^2\sigma^2}{2L^2}$ term). The latter arises because a fraction $\theta = \Lambda/L$ is invested in the firm, which has volatility σ.

We revisit Section 2.2. The least talented CEO (number N) has a reservation wage $\underline{w}_N$. To compensate for the above utility loss, he must be paid $\underline{w}_N e^\chi$. Hence the pay of CEO n is the following variant of equation (41):

$$w\left(n\right) = -\int_n^N CS\left(u\right)^\gamma T'\left(u\right)du + \underline{w}_N e^\chi \tag{30}$$

and scales according to

$$w\left(n\right) = D\left(n_*\right)S(n_*)^{\beta/\alpha}\left(S\left(n\right)^{\gamma-\beta/\alpha} - S\left(N\right)^{\gamma-\beta/\alpha}\right) + \underline{w}_N e^\chi \sim D\left(n_*\right)S(n_*)^{\beta/\alpha}S\left(n\right)^{\gamma-\beta/\alpha}$$

Changes in χ have very little effect on the the pay of top CEOs. Equation (30) shows that the pay of CEO n is composed of the rent to talent (the first term) and the wage of the least talented CEO (the second term). An increase in χ affects only the wage of the least talented CEO, and does not affect the rent to talent. Since the first term is much larger, particularly for highly talented CEOs, the overall wage is barely affected.

Proposition 8 *(Optimal contract in the extended model, general equilibrium). Let n_* denote the index of a reference firm. In equilibrium, the manager of rank n runs a firm of rank n, and receives an expected pay: $w = D\left(n_*\right)CS(n_*)^{\beta/\alpha}S^{\gamma-\beta/\alpha}$, where $D\left(n_*\right) = -n_*T'\left(n_*\right)/\left(\alpha\gamma - \beta\right)$. The wealth-performance sensitivity of a firm is as before, $\theta = \partial\ln W/\partial r = \Lambda/L$, and the scaling with size are the same as in the basic model of Section 2.*

4.2.2 Firms Differ in Parameters Besides Size

We now study the case of heterogeneity in the firm's cost of effort, scope of effort and volatility.[15] Let

$$\chi_n = \Lambda_n\bar{e}_n + \frac{\Gamma\Lambda_n^2\sigma_n^2}{2L_n^2} \tag{31}$$

---

[15]Cross-sectional variation in $\bar{e}_n$ reflects the fact that there is greater scope to add value through effort in certain companies and industries (e.g. those intensive in human capital).

denote the equivalent variation associated with firm n. If a CEO is hired by firm n for a nominal wage w, after adjustment for the cost of effort and risk aversion, the "effective" wage is $\nu = w_n e^{-\chi_n}$. Assume that in market equilibrium, a CEO of talent m receives an effective wage $v(m)$. If firm n wishes to hire manager m, it must pay him a net wage $v(m)$, and a dollar wage $v(n) e^{\chi_m}$. So its program is: $\max_m CS(n)^\gamma T(m) - v(n) e^{\chi_n}$, i.e.

$$\max_m C_n e^{-\chi_n} S(n)^\gamma T(m) - v(n)$$

Firm n behaves like a firm with "effective size" $(C_n e^{-\chi_n})^{1/\gamma} S(n)$. We next apply Gabaix and Landier (2008, Proposition 3), and assume that the firms' χ's are drawn independently of firm size. We obtain the following.

Proposition 9 *(Optimal contract in the extended model, general equilibrium, heterogeneous firms. Let n_* denote the index of a reference firm. In equilibrium, the manager of rank n runs a firm whose "effective size" $(C_i e^{-\chi_i})^{1/\gamma} S$ is ranked n, and receives an expected pay:*

$$w = D(n_*) C S(n_*)^{\beta/\alpha} S^{\gamma - \beta/\alpha} \exp\left(\frac{\beta}{\alpha\gamma} (\chi - \overline{\chi})\right), \tag{32}$$

where $D(n_*) = -n_* T'(n_*) / (\alpha\gamma - \beta)$ and $\overline{\chi}$ is the following average over the firms' equivalent variations $\widetilde{\chi}$:

$$e^{-\overline{\chi}} = E\left[e^{-\widetilde{\chi}/(\alpha\gamma)}\right]^{\alpha\gamma}. \tag{33}$$

The wealth-performance sensitivity of firm i is as before, $\theta_i = \partial \ln W / \partial r_i = \Lambda_i / L_i$, and the scaling with size are the same as in the basic model of section 2.

To interpret the Proposition, first note that the equivalent variation (31) χ_n increases in the cost of effort required by the firm ($\Lambda_n \overline{e}_n$), the risk of the firm (σ_n), and the required sensitivity of incentives (Λ_n / L_n). A firm with higher equivalent variation χ will, ceteris paribus, choose a lower quality manager (since its effective size is $S e^{-\chi/\gamma}$), but with a higher pay. This is because the effective size $S e^{-\chi/\gamma}$ leads to a net wage $v \propto \left(S e^{-\chi/\gamma}\right)^{\gamma - \beta/\alpha}$, and a full wage $w = v e^{\chi} \propto S^{\gamma - \beta/\alpha} e^{\chi\beta/\alpha\gamma}$, which is increasing in χ.

Hence in the cross-section, firms with high equivalent variations pay more. However, in the aggregate, there is no such effect: if the equivalent variation of all firms increases by the same amount δ, the wages do not change. This was demonstrated in the previous subsection, and here arises because both χ and $\overline{\chi}$ increase by δ, which creates no change in wage in (32).

5   Extensions and Alternative Specifications

This section considers extensions and other specifications of the one-period model. Section 5.1 shows that the multiplicative functional forms we used are necessary and sufficient to explain

the size-independence of B^I found in the data, since additive specifications do not generate the same prediction. Section 5.2 examines a second feature of traditional models, the assumption of an unbounded effort domain. Our model features bounded effort and has different predictions for the relationship between firm volatility and wealth volatility, which we support empirically. Section 5.3 considers actions that are additive in firm value, such as perk consumption. Section 5.4 reconciles our results with the empirical results of Baker and Hall (2004).

5.1   The Requirement for Multiplicative Preferences

With (1) we used preferences that are multiplicative in consumption and (a function of) effort: $E\left[cg\left(e\right)\right]$. Preferences such as $E\left[\phi\left(cg\left(e\right)\right)\right]$ would work the same way. This is a sufficient condition for b^I to be independent of w, which we have shown empirically. We now demonstrate the necessity of multiplicative preferences for generating this prediction.

Many previous theories of CEO pay (Haubrich (1994), Schaefer (1998), Baker and Hall (2004)) are based on the classical "additive" model of Holmstrom and Milgrom (1987), of the form $E\left[c\right]-g\left(e\right)$. We explore the implications of this specification while maintaining the same contract structure (equation (3)). We normalize the expected return to 0, and call b the fraction of w invested in stock, so that $c=w\left(1+br\right)$. We note that b is also $b^I \equiv E\left[\partial c/\partial r\right]/E\left[c\right]$. With the utility function $E\left[c\right]-g\left(e\right)$, the optimal b^I is given by $b^I=\frac{g(-1)-g(0)}{Lw}$, which implies:[16]

$$b^I \propto w^{-1} \tag{34}$$

The additive form therefore predicts that b^I decreases with the wage. This contrasts with the multiplicative form (1), which predicts that b^I is independent of the wage and is thus consistent with the data.

Another popular utility function is $E\left[c^\alpha/\alpha\right]-g\left(e\right)$, with $\alpha \in \left(0,1\right]$. This leads to $b^I \propto w^{-\alpha}$ for large w, and thus also predicts that b^I declines with firm size. The reason is that, for sufficiently high consumption, effort has a very small effect on the agent's utility and so fewer incentives are required to ensure compatibility.

While the above considered two specific functional forms, we now demonstrate a general result: that multiplicative preferences are necessary to generate a size-independent b^I. To keep the analysis streamlined, we consider only a highly simplified setup. Consider a general utility function is $E[u\left(c,e\right)]$, with $e \in \{-1,0\}$. Assume the firm's return is $r=Le$ and that incentive compensation is implemented with shares, so the firm selects expected pay $\bar{c}$ and slope b so that: $c=\bar{c}\left(1+br\right)$. The optimal contract minimizes $\bar{c}$ and b while granting the CEO his reservation utility of $\underline{u}$ and eliciting $e=0$.[17] The next Proposition states that multiplicative preferences

---

[16]The proof is as follows. The optimal b^I is the smallest b such that $E\left[c-g\left(0\right) \mid e=0\right] \geq E\left[c-g\left(-1\right) \mid e=-1\right]$, and so satisfies $E\left[c-g\left(0\right) \mid e=0\right] = E\left[c-g\left(-1\right) \mid e=-1\right]$. Since $c=w\left(1+br\right)=w\left(1+b\left(Le+\eta\right)\right)$, the conditions read: $w-g\left(0\right)=w\left(1-bL\right)-g\left(-1\right)$, i.e. $b=\frac{g(-1)-g(0)}{Lw}$.

[17]More fully, $\underline{v}=E\left[v\left(c,e\right) \mid e=0\right] \geq E\left[v\left(c,e\right) \mid e=-1\right]$.

are required for the optimal $b = E\left[\partial c / \partial r\right] / E\left[c\right]$ to be independent of $\underline{v}$ (and thus $E[c]$).

Proposition 10 *(Necessity and sufficiency of multiplicative preferences to generate a size-independent b^I). Assume the CEO's utility function is $u\left(c, e\right)$, with c consumption and e effort, and the firm's return is $r = Le$. Suppose the optimal affine contract involves a pay scaled pay-performance sensitivity $b^I = E\left[\partial c / \partial r\right] / E\left[c\right]$ that is is independent of $E\left[c\right]$. Then, the utility function is multiplicative in consumption and effort, i.e. can be written:*

$$u\left(c, e\right) = \phi\left(c \cdot g\left(e\right)\right) \tag{35}$$

for some functions ϕ and g.

Conversely, if preferences are of the type (35), then the optimal contract has a slope b that is independent of $E\left[c\right]$.

We note that the above Proposition was proven in a restrictive context, with no noise and restricting the contract to consist of cash and shares, although we considered a general utility function. We suspect that the results extend to more general settings, but such an investigation is beyond the central objective of this paper.[18]

5.2 Bounded Effort and the Link Between Wealth Volatility and Firm Volatility

A second feature of the traditional additive model is that it features unbounded effort. We show that this assumption leads to a predicted negative association between pay volatility and firm volatility. This contrasts both our model and the data.

Under the additive (exponential-normal) model, the CEO has utility $u = E[c] - \frac{a}{2} var\left(c\right) - \frac{1}{2}e^2$, where a denotes absolute risk aversion and $e \in [0, \infty)$. His reservation utility is $\underline{u}$. Firm value next period is $S_1 = S\left(1 + \mu + Le + \eta\right)$, where L measures the CEO's productivity, and η is stochastic noise with mean 0 and variance σ_r^2. μ accounts for the firm's expected returns in equilibrium. The firm maximizes $S\left(1 + \mu + Le\right) - E\left[c\right]$, its expected value next period net of CEO pay. As before, compensation comprises fixed pay f, plus ν shares.

The solution is standard.[19] The CEO's dollar-dollar-pay-performance sensitivity is $b^{II} = \partial c / \partial S_1 = L / \left(L^2 + a\sigma_r^2\right)$, and thus is decreasing in firm volatility. This well-known prediction stems from the fact that there is always an interior solution to the optimal effort level, and so it reflects a trade-off between risk and incentives at the margin. As σ_r rises, the trade-off

---

[18]For instance, with noise, we suspect that to keep b constant across expected utilities, the function ϕ must actually be: $\phi\left(c\right) = A \ln c + B$ or $Ac^{1-\Gamma} / \left(1 - \Gamma\right) + B$.

[19]Normalizing the initial share price to $P = 1$, the CEO's realized pay is $c = f + \nu\left(1 + Le + \eta\right)$. The CEO chooses e to maximize his utility, $U = f + \nu\left(1 + Le\right) - \frac{a}{2}\sigma_r^2\nu^2 - \frac{1}{2}e^2$, and selects $e = \nu L$. The firm chooses ν to maximize its net value, $S\left(1 + \nu L^2\right) - \frac{a}{2}\sigma_r^2\nu^2 - \frac{\nu^2 L^2}{2}$, and selects $\nu = SL^2 / \left(L^2 + a\sigma_r^2\right)$. The CEO's total pay is therefore $c = f + S_1 L / \left(L^2 + a\sigma_r^2\right)$.

leads to optimal incentives being lower. By contrast, our model predicts that pay-performance sensitivity is independent of firm size (see Section 2). The evidence from Prendergast (2002) finds little evidence of a negative relationship between incentives and firm volatility.

In addition, models with bounded effort predict a negative relationship between pay volatility and firm volatility. Since pay volatility is $stdev\left(c\right) = \nu\sigma_r = \sigma_r SL/\left(L^2 + a\sigma_r^2\right)$, its sensitivity to firm volatility is given by $\partial stdev\left(c\right)/\partial\sigma_r = -S\left(1 - 2b^{II}\right)b^{II}$. Since empirical studies find that b^{II} is substantially less than $1/2$, these models predict

$$\partial stdev\left(c\right)/\partial\sigma_r < 0.$$

By contrast, in our model there is a corner solution to effort and so the number of shares ν is independent of volatility. Hence $stdev\left(c\right) = \nu\sigma_r$ is increasing in volatility. Indeed, we predict that the CEO's wealth volatility is proportional to firm volatility, i.e.

$$stdev(W_{t+1} - W_t) = B^{III}\sigma_r \propto S^\rho\sigma_r, \tag{36}$$

where σ_r is the volatility of the firm's returns and $\rho = 1/3$ is the elasticity of pay with respect to size (see Proposition 4).

We now evaluate these contrasting predictions using the same dataset as before.[20] As discussed more fully in Appendix A, there are two main ways to estimate wealth volatility, $stdev(W_{t+1} - W_t)$. The first is the ex ante measure used in Section 3, i.e. $stdev(W_{t+1} - W_t) = B_t^{III}\sigma_r$.[21] The second uses ex post realized volatility, i.e. $stdev(W_{t+1} - W_t) = \ln|W_{t+1} - W_t|$. In both cases, the model predicts that regressing $stdev(W_{t+1} - W_t) = \beta_S \ln S + \beta_S \ln\sigma_r$, will yield $\beta_S = 1/3$ and $\beta_\sigma = 1$.

We can also scale the dependent variable. Scaling by the wage leads to $B_t^I\sigma_r$ or $\ln\left(|W_{t+1} - W_t|/w_t\right)$ and the model predicts $\beta_S = 0$ and $\beta_\sigma = 1$. Scaling by size yields $B_t^{II}\sigma_r$ or $\ln\left(|W_{t+1} - W_t|/S_t\right)$, with a prediction of $\beta_S = -2/3$ and $\beta_\sigma = 1$.

The results are shown in Table 3. In all six specifications we find that wealth volatility is significantly positively linked to firm volatility. In three specifications, we cannot reject the hypothesis that $\beta_\sigma = 1$. (The low $\beta_\sigma = 0.64$ when $\ln\left(|W_{t+1} - W_t|/w_t\right)$ is the dependent variable is because of the strong positive association between w_t and σ_r.) In addition, in all six specifications, the 95% confidence intervals for β_S contain the predicted values. In unreported regressions we find that these results are unchanged when adding firm fixed effects and identifying purely on within-firm changes in volatility.

Insert Table 3 about here

---

[20]The linear-quadratic model is expressed in terms of terminal consumption, but its general meaning is in terms of terminal wealth. The key variable is the NPV of the CEO's future utilities in the second period, which is also linear in wealth in the linear-quadratic model.

[21]Indeed, for small time intervals, $W_{t+1} - W_t = W_t'\left(r\right)r_t = B^{III}r_t$, so $stdev(W_{t+1} - W_t) = B^{III}stdev\left(r_t\right) = B^{III}\sigma_r$.

5.3 Perks

In the basic model, where the contract consists of cash and shares, the analysis assumed that $L > \Lambda$, and thus incentive problems were solvable through the contract specified in Proposition 1. However, if the assumption is violated, the manager's disutility from working is so high that a large equity stake is needed to induce the correct action. If expected pay is kept at w, this necessitates a negative fixed component f, which violates limited liability. One important agency problem for which $L < \Lambda$ might apply is CEO entrenchment. Since resigning adversely impacts the CEO in future periods, the total loss in utility is likely far greater than from exerting greater effort or forgoing an empire-building merger. Since incentive pay is ineffective at inducing underperforming CEOs to leave, this issue must instead be addressed by corporate governance, such as active boards. (This solution is also not unproblematic since boards may be endogenously chosen by the CEO, as modeled by Hermalin and Weisbach (1998)).

Moreover, the necessary condition for incentive pay to be effective is substantially stronger if the effort decision is additive in firm value. This is likely the case for perks, such as corporate jets: the value loss from perk consumption is relatively independent of firm size.

Proposition 11 (*Impossibility of deterring perk consumption through incentive pay*). *Assume $e = -1$ reduces firm value by \$$\overline{L}$, i.e. $\overline{L} = SL$ in the prior analysis. Let $\overline{L} > w\Lambda$, so that $e = 0$ maximizes total surplus. It is impossible to elicit high effort while keeping expected pay fixed at w if $S > \overline{L}/\Lambda$, i.e. the firm is sufficiently large.*

Hence if $w\Lambda < \overline{L} < S\Lambda$, perk consumption is inefficient but cannot be prevented. Since the perk is fixed in absolute terms, the stock price of a large firm is relatively insensitive to perk consumption. Therefore, the CEO's equity stake does not decline sufficiently in dollar terms to outweigh the utility gain of perk consumption. Note that perks cannot be prevented even if the firm is willing to pay the CEO rents, i.e. a pay in excess of $w(n)$, by awarding him a large number of shares. Raising the CEO's pay augments his utility from perk consumption (as this equals $w\left(1 - \frac{1}{\Lambda}\right)$) so incentive compatibility is still not achieved. The only possible solution would be to give the CEO a large equity stake and reduce his fixed salary, to keep his total pay constant, but this is not possible as $f \geq 0$.

While seemingly intuitive, this result is contrary to the view modeled by Jensen and Meckling (1976) and implied by empirical papers such as Jensen and Murphy (1990), that agency costs can (and should) be addressed by incentive pay. Equity compensation is primarily effective in addressing agency costs that are a proportion of firm value, such as effort or M&A.[22] However, perks are typically independent of firm value, and thus have very little explanatory power for observe incentives. As with the entrenchment issue, perks should be controlled by active corporate governance. For example, the board could intensely scrutinize the purchase of a

---

[22]For example, Morck, Shleifer and Vishny (1990) find that higher managerial equity stakes are associated with greater value creation in mergers.

corporate jet or a large investment project. Empirical evidence linking governance to firm performance (e.g. Gompers, Ishii and Metrick (2003) and Yermack (2006)) can be interpreted as consistent with this result. If all agency costs could be solved by incentive compensation, governance would not matter (except for ensuring that the CEO is given the optimal contract). Since incentive compensation is not universally effective, there remains a substantial incremental role for governance.

Compensation continues to be ineffective at deterring perk consumption even when allowing for more general incentive contracts. The intuition remains the same: since perks have a very small effect on the return of a large firm, incentive compatibility requires extremely high wealth-performance sensitivity. Even if this can be achieved by a general contract (e.g. an option), the risk it imposes on the CEO would outweigh the gains from deterring the perk. Appendix C.2 formalizes this point.

5.4 Explaining Baker-Hall

Finally, we illustrate how our model can explain Baker and Hall's (2004) empirical results on the negative relationship between B^{II} and firm size. They assume an additive model, which requires L to be size-dependent in order to predict that B^{II} scales with size. They therefore use their results to calibrate the scaling of L with size. We show that their findings are also consistent with our model, in which L is constant and size-dependence is instead generated by the multiplicative functional form.

Using our notation, Baker and Hall estimate a functional form for $L(e, S)$. They derive an equation for CEO productivity as a function of firm size: $I^{BH} = \sqrt{\frac{2b^{II}a}{1-b^{II}}}\sigma_r S$ (their equation (3)), where a is the coefficient of absolute risk aversion.[23] They assume constant relative risk aversion, and so a is inversely proportional to the CEO's wealth.

They then make one of three assumptions for the scaling of the CEO's wealth, which leads to three different specifications. In their specification (1), they assume wealth is proportional to the CEO's wage, and so $a \propto w^{-1}$. In our model, $w \propto S^\rho$ and so $a \propto 1/w \propto S^{-\rho}$. In addition, $b^{II} \propto w/S \propto S^{\rho-1}$ and $1 - b^{II} \propto S^0$, since $b^{II} \ll 1$. Assuming stock price volatility is independent of firm size (as in the geometric random growth model),[24] the standard deviation of the dollar value of a firm is $\sigma_r \propto S^1$. We therefore predict: $I_1^{BH} \propto S^{(\rho-1-\rho)/2+1} = S^{1/2}$. Our predicted elasticity of $\frac{1}{2}$ is consistent with Baker and Hall's empirical finding of 0.4.

In their specification (3), they assume the CEO's wealth is independent of size, and therefore $a \propto S^0$. In our model, this would lead to: $I_3^{BH} \propto S^{(\rho-1)/2+1} = S^{(1+\rho)/2} = S^{2/3}$, using $\rho = 1/3$,

---

[23]Baker and Hall (2004) use ρ to denote absolute risk aversion; we are using a to avoid confusion with our ρ, which denotes the elasticity of total pay with respect to firm size. Also, we use σ to note the "percentage" volatility of the firm.

[24]Regressing log volatility on log aggregate value, year dummies and industry dummies yields an insignificant coefficient of -0.0024 (standard error of 0.0119).

and thus a predicted elasticity of 0.67. Baker and Hall find an elasticity of 0.62. We therefore conclude that the Baker and Hall results can also be explained quantitatively by our framework.

6 Conclusion

The primary contribution of this paper is to develop a calibratable equilibrium model of both the total level and incentive component of CEO compensation. The model can be used to evaluate a number of ongoing debates on the efficiency of compensation practices. Since Jensen and Murphy (1990), a number of researchers have documented a strong negative relationship between dollar-dollar incentives and size. One common interpretation is that contracts are especially suboptimal in large firms, perhaps resulting from the CEO's excessive influence on his own pay.

This paper has a different conclusion. The predictions generated by our neoclassical benchmark closely match the data, implying that the widely documented empirical scalings are fully consistent with optimal contracting. It is indeed efficient for effective equity stakes to be particularly low in large firms, and for dollar holdings to rise less than proportionately with firm size. Similarly, the model is able to explain the level of CEO incentives without appealing to rent extraction.

We discuss a number of other applications of the model. We demonstrate, both theoretically and empirically, that scaled wealth-performance sensitivity (the dollar change in CEO wealth for a percentage change in firm value, divided by annual pay) is stable across firm size. This property renders it particularly attractive in a number of empirical applications, and is not shared by other measures of incentives previously used.

A further application is to understand additional determinants of compensation, over and above the factors considered in Gabaix and Landier (2008), generating predictions that could be investigated in future empirical research. The model suggests that cross-sectional differences in risk and effort lead to variation in wages between firms, but aggregate-level increases in these variables have negligible effect. (This is potentially testable between firms, between industries or between countries). In addition, total salary should be independent of wealth-performance sensitivity: the former is determined by "pay-for-talent", not "pay-for-performance". There are additional empirical predictions from the core model that have not been tested in the paper. Are our scalings empirically consistent in other countries, or are there large discrepancies that may be potential evidence of inefficiencies? Are CEO incentives increasing in wealth?[25] How much of the time series variation in incentives, documented by Frydman and Saks (2007) and Jensen and Murphy (2004), can be explained by our model?

---

[25] Given data limitations in the U.S., the only wealth data available is on the CEO's stock and options holdings in his own firm, and so there is a mechanical link between incentives and measured wealth. However, full wealth data may be available in other countries (see Becker (2006) for an example).

Over and above these applications, the model itself may be of use for future researchers in executive compensation. It is a simple, market equilibrium framework that matches the following "trilogy of scalings", a potential criterion for the empirical accuracy of future models: the 1/3 elasticity of pay with respect to firm size, the independence of %-% incentives to size, and the proportionality between wealth volatility and firm volatility. As such, it may form a building block upon which more complex models can be built. In particular, there are a large number of additional complexities in the real world upon which the model is silent, and it would be interesting to quantify their equilibrium implications and investigate whether they can explain other observed features of compensation. Examples include accounting performance measures (which may explain bonuses), entrenchment and turnover (which may explain severance pay), stockholder-bondholder conflicts (which may explain inside debt compensation, such as pensions), and renegotiation.

Our conclusions should be tempered by a number of observations. First, our model's prediction that B^I is size invariant stemmed from our assumed functional forms, and other specifications would have different predictions. We used the quantitative empirical consistency of our model to justify our assumptions and thus our advocacy of B^I as an empirical measure. However, using real-world data to evaluate a frictionless model implicitly assumes that real-world practices are also reasonably close to frictionless. It could be that an alternative model, with different specifications to ours and predicting the size invariance of a different measure, represents the "true" frictionless benchmark, and that this alternative model is empirically rejected because there are indeed inefficiencies in reality. Perhaps under the hypothetical "true" specification, B^I should optimally increase with firm size, and we only observe that it is constant because inefficiencies are greater in large firms. Further research is needed to evaluate this hypothesis. In particular, the strongest support for the rent extraction view may come not from observing that a particular practice is inconsistent with a frictionless model, but from deriving a model that explicitly incorporates frictions and generates quantitative predictions on their effects on compensation that closely match the data. Our empirical results suggest that, if the "true" specification predicts that B^I increases with firm size, inefficiencies would have to scale with firm size in such a way as to exactly counterbalance the optimal scaling and explain the size invariance of B^I that we find. For now, our neoclassical benchmark shows that inefficiencies do not need to be assumed to be able to match various features of the data.

Second, in our model, incentive compensation can induce the CEO to take the first-best action. However, certain agency problems cannot be solved through pay: perks (if the firm is large), entrenchment, and effort decisions where the CEO's disutility is sufficiently high. This leads to an ongoing role for corporate governance, over and above the selection of the optimal contract.

A Detailed Calculation of B^I

We merge Compustat with ExecuComp (1992-2005) and each year select the 500 largest firms by aggregate value (equity plus debt). To calculate aggregate value, we first multiply the end-of-year share price (data199) with the number of shares outstanding (data25) to obtain market equity. To this we add the value of the firm's debt, calculated as total assets (data) minus total common equity (data60) and minus balance sheet deferred taxes (data74). We call this variable aggval, and it is in millions of dollars.

The CEO's incentives are calculated at the end of each fiscal year, and stem from his stock and option holdings. The number of shares held by the CEO is given by ExecuComp variable shrown. Obviously, each share has a delta of 1; the delta of an option is given by the Black-Scholes formula:

$$e^{-dT} N \left(\frac{\ln \left(\frac{S}{X} \right) + \left(r - d + \frac{\sigma^2}{2} \right) T}{\sigma \sqrt{T}} \right).$$

d is the continuously compounded expected dividend yield, given by bs_yield. If this is missing, we assume it is zero. We also winsorize it at the 95th percentile for each year.

σ is the expected volatility of the stock return, given by bs_volat. If it is missing, we replace it with the mean volatility for that year, given by `http://mi.compustat.com/docs-mi/help/blk_schol.htm`. We also winsorize σ at the 5th and 95th percentile for each year.

r is the continuously compounded risk-free rate, available from `http://mi.compustat.com/docs-mi/help/blk_schol.htm`.

S is the stock price at the end of the fiscal year, given by prccf.

X is the strike price of the option.

T is the maturity of the option.

The option holdings come in three categories: new grants, existing unexercisable grants, and existing exercisable grants. The first four variables in the Black-Scholes formula are available for all categories. For new grants, X and T are also available. X is given by expric, and T can be calculated using the option's maturity date, exdate. If exdate is unavailable, we assume maturity of 10 years. A CEO may receive multiple new grants in each year. We calculate the delta of each option grant, multiply it by the number of options in the grant (numsecur) and sum across grants to calculate "totaldeltanew", the dollar change in the CEO's newly granted options for a $1 increase in the stock price. Similarly, we sum numsecur across grants to calculate "numnewop", the total number of newly granted options. While ExecuComp has a variable (soptgrnt) for the number of newly granted options, it is sometimes different from the number obtained by summing across grants. As will become clear later, using the "bottom-up" number numnewop is more internally consistent since we are calculating the intrinsic value of

new grants on a "bottom-up" basis.

X and T are not directly available for previously granted options, so we use the methodology of Core and Guay (2002a). Here we summarize the Core and Guay method while stating the additional assumptions made when data issues were encountered. Since new grants are nearly always unexercisable, Core and Guay recommend calculating the strike price of unexercisable options as

$$\text{prccf} - \frac{\text{inmonun - ivnew}}{\text{uexnumun - numnewop}}.$$

inmonun is the intrinsic value of the unexercisable options held at the end of the year, some of which stem from newly granted options.

ivnew is the intrinsic value of the newly granted options. This is not directly available from ExecuComp, but obtained by calculating max(0,(prccf-expric)) * numsecur for each new grant and summing across new grants.

uexnumun is the number of unexercisable options held at the end of the year.

Again because new grants are nearly always unexercisable, Core and Guay recommend calculating the strike price of exercisable options as

$$\text{prccf} - \frac{\text{inmonex}}{\text{uexnumex}}.$$

inmonex is the intrinsic value of the exercisable options held at the end of the year.

uexnumex is the number of exercisable options held at the end of the year.

In some cases, numnewop > uexnumun, i.e. the number of newly granted options exceeds the number of unexercisable options at year end. We interpret these cases as part of the new grant (numnewop - uexnumun) being exercisable. We therefore calculate the strike price of exercisable options as

In a subset of these cases, numnewop > uexnumun + uexnumex, i.e. the number of newly granted options exceeds the number of total options at year end. In such cases, we assume that the options held at year end entirely stem from new grants and there were no previously granted options.

In some cases, ivnew > inmonun, i.e. the intrinsic value of the newly granted options exceeds the number of unexercisable options. In a subset of these cases, uexnumun > numnewop, i.e. there are some previously granted unexercisable options, and their deltas need to be taken into account. We assume that such options are at the money. If ivnew > inmonun and numnewop > uexnumun, we interpret this as part of the new grant being exercisable and having intrinsic value. In such cases, we calculate the strike price of exercisable options as

$$\text{prccf} - \frac{\text{inmonex - (ivnew - inmonun)}}{\text{uexnumex - (numnewop - uexnumun)}}.$$

If ivnew > inmonex + inmonun but uexnumex > numnewop - uexnumun, i.e. there are some previously granted exercisable options, and their deltas need to be taken into account, we assume that these options are at the money.

For the option maturities, Core and Guay recommend assuming a maturity for previously granted, unexercisable options of one year less than the maturity of newly granted options, if there were new grants in the fiscal year. (Where there were multiple grants, we take the longest maturity option). If there were no grants, Core and Guay recommend a maturity of 9 years. The maturity of exercisable options is assumed to be 3 years less than for unexercisable options. If this leads to a negative maturity, we assume a maturity of 1 day. As in Core, Guay and Verrecchia (2003), we then multiply the maturities of all options by 70%, to capture the fact that CEOs typically exercise options prior to maturity.

We use these estimated strike prices and maturities to calculate "deltaun", the delta for previously granted, unexercisable options, and "deltaex", the delta for previously granted, exercisable options.

Putting this all together, the dollar change (in millions) in the CEO's wealth for a $1 change in the stock price is given by

$$
\begin{aligned}
\text{totaldelta} = [\quad & \text{shrown} + \text{totaldeltanew} + \max(0,\text{uexnumun-numnewop}) \times \text{deltaun} \\
+ \quad & \max(0,(\text{uexnumex-}\max(0,\text{numnewop-uexnumun}))) \times \text{deltaex}]/1000.
\end{aligned}
$$

We then calculate our measures of wealth-performance sensitivity:

$$
\begin{aligned}
B^{III} &= \text{totaldelta} \times \text{prccf} \\
B^{II} &= \frac{B^{III}}{\text{aggval}} \times 1000 \\
B^{I} &= \frac{B^{III}}{\text{tdc1}} \times 1000.
\end{aligned}
$$

Since tdc1 is very low (and sometimes zero) in a few observations, we replace such observations by the 2nd percentile for that year. The units for B^{II} are the dollar increase in the CEO's wealth for a $1,000 dollar increase in shareholder value, as in Jensen and Murphy (1990).

Note that these "ex ante" measures slightly underestimate wealth-performance sensitivity, since they omits changes in flow compensation. However, this discrepancy is likely to be small: Hall and Liebman (1998) and Core, Guay and Verrecchia (2003) find that the bulk of incentives comes from changes in the value of a CEO's existing portfolio. If the researcher has data on the CEO's entire wealth, B^{I} can be estimated using ex post changes in wealth as follows:

$$
\frac{W_{t+1} - W_t}{w_t} = A + \widehat{B^I} \times r_{t+1} + C \times r_{M,t+1} + \text{Controls}, \tag{37}
$$

where $W_{t+1} - W_t$ is the change in wealth and $r_{M,t+1}$ is the market return (returns on other factors could also be added).[26] This compares with our chosen measure of:

$$B^{I,\text{ex ante}} = \frac{1}{w_t} \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P \right], \tag{38}$$

where V is the value of one option, $\frac{\partial V}{\partial P}$ is the option "delta", and P is the stock price.

Even if full wealth data (which includes flow compensation) is available, the ex ante measure has a number of advantages. First, both data on overall wealth and a long time series are required to estimate equation (37) accurately. Second, even if such data is available, ex post measures inevitably assume that wealth-performance sensitivity is constant over the time period used to calculate the measure. Since the ex ante statistic more accurately captures the CEO's incentives at a particular point in time, it is especially useful as a regressor since its time period can be made consistent with the dependent variable. For example, in a regression of M&A announcement returns on wealth-performance sensitivity (e.g. Morck, Shleifer and Vishny (1990)), the CEO's incentives can be measured in the same year in which the transaction was announced. In a similar vein, the ex ante measure is more suited to measuring trends in executive compensation over time.

Finally, if the researcher only has data on compensation flows, rather than wealth, this typically significantly understates wealth-performance sensitivity. However, if the CEO is known to have limited shares and options, the pay-performance estimate b^I will be a reasonable approximation:

$$\ln w_{t+1} - \ln w_t = a + \widehat{b^I} \times r_{t+1} + \text{Controls}, \tag{39}$$

where w_t is flow compensation and r_t is the firm's return. Variations on the above specification are possible. For example, an alternative dependent variable is $2 (w_{t+1} - w_t) / (w_{t+1} + w_t)$, which is more robust when w_t is close to 0.

B Detailed Proofs

Proof of Proposition 1 The manager should earn his market wage: $E[c \mid e = 0] = w$. We calculate:

$$
\begin{aligned}
E[c \mid e = 0] &= f + \nu P = w \\
E[c \mid e = -1] &= f + \nu P (1 - L) = f + \nu P - \nu P L = w - \nu P L.
\end{aligned}
$$

---

[26] $r_{M,t+1}$ is added since the CEO may hold investments other than his own firm's securities, that move with the market but not the firm's return. For example, consider a CEO whose wealth is entirely invested in the market, with no sensitivity to firm's idiosyncratic return. If equation (37) did not contain the $C \times r_{M,t+1}$ term, it would incorrectly find $\widehat{B^I} > 0$, whereas the true $\widehat{B^I}$ is zero. Since r_{t+1} proxies for $r_{M,t+1}$, there is an omitted variables bias which leads to B^I being overestimated.

The manager chooses $e = 0$ if:

$$E\left[cg\left(0\right) \mid e = 0\right] \geq E\left[cg\left(-1\right) \mid e = -1\right].$$

Since $g\left(0\right) = 1$ and $g\left(-1\right) = \frac{1}{1-\Lambda}$, this implies

$$w \geq \frac{w - \nu PL}{1 - \Lambda} \Leftrightarrow \nu P \geq \nu^* P = w\frac{\Lambda}{L}.$$

f^* is chosen to ensure that expected pay is w: $^* = w - \nu^* P = w\left(1 - \frac{\Lambda}{L}\right)$.

Proof of Proposition 2 We first define some notation. A continuum of firms and potential managers are matched together. Firm $n \in [0, N]$ has size $S\left(n\right)$ and manager $m \in [0, N]$ has talent $T\left(m\right)$. Low n denotes a larger firm and low m a more talented manager: $S'\left(n\right) < 0$, $T'\left(m\right) < 0$. $n\left(m\right)$ can be thought of as the rank of the manager (firm), or a number proportional to it, such as its quantile of rank.

We consider the problem faced by one particular firm. The firm has a "baseline" value of S. At $t = 0$, it hires a manager of talent T for one period. The manager's talent increases the firm's value according to

$$S' = S + CTS^\gamma, \tag{40}$$

where C parameterizes the productivity of talent. If large firms are more difficult to change than small firms, then $\gamma < 1$. If $\gamma = 1$, the model exhibits constant returns to scale (CRS) with respect to firm size.

We now determine equilibrium wages, which requires us to allocate one CEO to each firm. Let $w\left(m\right)$ denote the equilibrium compensation of a CEO with index m. Firm n, taking the market compensation of CEOs as given, selects manager m to maximize its value net of wages:

$$\max_m CS\left(n\right)^\gamma T\left(m\right) - w\left(m\right).$$

The competitive equilibrium involves positive assortative matching, i.e. $m = n$, and so $w'\left(n\right) = CS\left(n\right)^\gamma T'\left(n\right)$. Let $\underline{w}_N$ denote the reservation wage of the least talented CEO $\left(n = N\right)$. Hence we obtain the classic assignment equation (Sattinger (1993), Tervio (2007)):

$$w\left(n\right) = -\int_n^N CS\left(u\right)^\gamma T'\left(u\right) du + \underline{w}_N. \tag{41}$$

Specific functional forms are required to proceed further. We assume a Pareto firm size distribution with exponent $1/\alpha$: $S\left(n\right) = An^{-\alpha}$. Using results from extreme value theory, GL use the following asymptotic value for the spacings of the talent distribution: $T'\left(n\right) = -Bn^{\beta-1}$. These functional forms give the wage equation in closed form, taking the limit as $n/N \to 0$:

$$w\left(n\right) = \int_n^N A^\gamma BC u^{-\alpha\gamma+\beta-1} du + \underline{w} = \frac{A^\gamma BC}{\alpha\gamma - \beta}\left[n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}\right] + \underline{w}_N \sim \frac{A^\gamma BC}{\alpha\gamma - \beta} n^{-(\alpha\gamma-\beta)}. \tag{42}$$

To interpret equation (42), we consider a reference firm, for instance firm number 250 – the median firm in the universe of the top 500 firms. Denote its index n_*, and its size $S(n_*)$. We obtain Proposition 2 from GL, which we repeat here. In equilibrium, manager n runs a firm of size $S(n)$, and is paid according to the "dual scaling" equation:

$$w\left(n\right) = D\left(n_*\right) S(n_*)^{\beta/\alpha} S\left(n\right)^{\gamma-\beta/\alpha},$$

where $S(n_*)$ is the size of the reference firm and $D\left(n_*\right) = -C n_* T'\left(n_*\right)/\left(\alpha\gamma - \beta\right)$ is a constant independent of firm size.[27]

Proof of Proposition 5 Take the definition of b^{II} and use $\rho = \gamma - \beta/\alpha$:

$$b^{II} = \frac{\Lambda}{L}\frac{w}{S} = \frac{\Lambda}{L}\frac{D\left(n_*\right) S(n_*)^{\beta/\alpha} S^{\gamma-\beta/\alpha}}{S\left(n\right)} \propto \frac{S^{\gamma-\beta/\alpha-1}}{S(n_*)^{-\beta/\alpha}} = \frac{S^{\rho-1}}{S(n_*)^{\rho-\gamma}} = S^{-(1-\rho)} S\left(n_*\right)^{\gamma-\rho}.$$

The expressions for b^I and b^{III} are similarly obtained.

Proof of Proposition 7 We define:

$$a_T = \exp\left[\left(r_f - \pi\theta + \Lambda\overline{e} + \Gamma\sigma^2\theta^2/2\right) T\right] \tag{43}$$

$$b_T = \left(r_f\left(1-\theta\right) + \left(\theta - \theta^2\right)\frac{\sigma^2}{2}\right) T. \tag{44}$$

We follow the techniques of dynamic contract theory, as in Sannikov (2006) and He (forth.). We normalize $L = 1$ and $\pi = \overline{e} = 0$ in the proof. We define the promised utility $\left(1 - \Gamma\right) U_t = E_t^P\left[W_T^{1-\Gamma}\right] > 0$, under the probability induced by the policy P that the CEO always exerts the maximum effort, $e_t = 0$. To understand its behavior, let us first calculate it under the proposed policy, noted $*$. Since, under this proposed policy, $\frac{dW_t}{W_t} = \theta\frac{dP_t}{P_t} + \left(1-\theta\right) r_f dt$, with $\theta = \Lambda$:

$$W_T^* = W_t^* \exp\left(\Lambda\sigma\left(z\left(T\right) - z\left(t\right)\right) - \Lambda^2\sigma^2\left(T-t\right)/2\right)$$

---

[27] The derivation is as follows. Since $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_* T'\left(n_*\right) = -Bn_*^{\beta}$, we can rewrite equation (42) as follows:

$$\begin{aligned}\left(\alpha\gamma - \beta\right) w\left(n\right) &= A^\gamma BC n^{-(\alpha\gamma-\beta)} = CBn_*^{\beta} \cdot \left(An_*^{-\alpha}\right)^{\beta/\alpha} \cdot \left(An^{-\alpha}\right)^{(\gamma-\beta/\alpha)}\\ &= -Cn_* T'\left(n_*\right) S(n_*)^{\beta/\alpha} S\left(n\right)^{\gamma-\beta/\alpha}.\end{aligned}$$

we have

$$dU_t^*/U_t^* = (1 - \Gamma) \Lambda \sigma dz_t \tag{45}$$

In general, the promised utility U_t is a martingale, so, by the martingale representation theorem, it can be written: $dU_t = (1 - \Gamma) U_t \theta_t \sigma dz_t$ for some adapted process θ_t. Hence we define the promised utility process to be:

$$dU_t = (1 - \Gamma) U_t \theta_t (dP_t/P_t - r_f dt) \tag{46}$$

First, we examine the incentive compatibility condition. If the CEO chooses an effort e_t, then $dP_t/P_t - r_f dt = e_t dt + \sigma dz_t$, and his full utility $U_T \exp\left(- (1 - \Gamma) \Lambda \int_0^T e_t dt\right)$ evolves as:

$$
\begin{aligned}
dU_t - (1 - \Gamma) U_t \Lambda e_t dt &= (1 - \Gamma) U_t (\theta_t (dP_t/P_t - r_f dt) - \Lambda e_t dt) \\
&= (1 - \Gamma) U_t [(\theta_t - \Lambda) e_t dt + \theta_t \sigma dz_t]
\end{aligned}
$$

Incentive compatibility is achieved (i.e. CEO sets e_t at the maximum level) if and only if $\theta_t - \Lambda \geq 0$.

We now verify that the policy is cost-minimizing. For a promised utility process U_t^* under the candidate optimal policy, define

$$b(u, t) = E\left[((1 - \Gamma) U_T^*)^{1/(1-\Gamma)} \mid U_t^* = u\right] = ((1 - \Gamma) u)^{1/(1-\Gamma)} \exp\left(\frac{\Gamma \Lambda^2 \sigma^2}{2} (T - t)\right).$$

Now, consider an alternative incentive compatible policy U_t. Individual rationality stipulates $U_0 \geq u_0 = U_0^*$. Define $G_t = b(U_t, t)$. Since $b(U_t^*, t)$ is a martingale, and since by (45), $var(dU_t^*) = U_t^{*2} ((1 - \Gamma) \Lambda \sigma)^2 dt/2$, we have: $b_t(t, u) + b_{UU}(t, u) u^2 ((1 - \Gamma) \Lambda \sigma)^2/2 = 0$ for any u. Therefore, for the general $G_t = b(U_t, t)$,

$$E_t dG_t = b_t + b_{UU} U^2 ((1 - \Gamma) \theta_t \sigma)^2/2 = b_{UU} U^2 ((1 - \Gamma) \sigma)^2/2 \cdot (\theta_t^2 - \Lambda^2).$$

However, owing to the incentive compatibility condition, $\theta_t \geq \Lambda$. Therefore, dG_t has non-negative drift, and has 0 drift if it is the candidate optimal policy. This implies:

$$E_0[b(U_T, T)] \geq b(U_0, 0) \geq b(u_0, 0) = b(U_0^*, 0) = E_0[b(U_T^*, T)].$$

Since $b(U_T, T) = ((1 - \Gamma) U_T)^{1/(1-\Gamma)} = c_T$, we have: $E_0[c_T] = E_0[b(U_T, T)] \geq E_0[b(U_T^*, T)] = E_0[c_T^*]$, which means any incentive compatible policy has an expected cost weakly greater than the candidate optimal one. That means that the candidate optimal policy is indeed cost-minimizing.

Finally, the various deterministic terms of the type $\exp(Ct)$ arise from Ito's lemma. In particular, $E_0\left[W_t^{1-\Gamma}\right]^{1/(1-\Gamma)} = W_0 \exp\left(r + \theta\pi - \Gamma\theta^2\sigma^2/2\right) t$, $E_0[W_t] = W_0 e^{(r+\theta\pi)t}$, and $E_0\left[(P_t/P_0)^\theta\right] =$

$$\exp \left(r + \theta\pi + \left(\theta^2 - \theta \right) \sigma^2/2 \right) t.$$

Proof of Proposition 9   As in Gabaix and Landier (2008, Section V.A, equation (25)), the firm will pay a wage associated with its effective size $\left(C_n e^{-\chi_n} \right)^{1/\gamma} S(n)$, namely: $v = D(n_*) \left(e^{-\overline{\chi}} S(n_*) \right)^{\beta/\alpha} \left(C_n e^{-\chi/\gamma} S \right)^{\gamma - \beta/\alpha}$. After the compensating differential, the dollar wage is: $w = v e^{\chi}$, hence (32).

Proof of Proposition 10   Define $\phi(c) = u(c, 0)$, $g(0) = 1$ and $g(-1) = 1/(1 - bL)$. Call $b = E[\partial c/\partial r]/E[c]$ the slope. Since b offers the minimum slope, $E[v(c, e) \mid e = 0] = E[u(c, e) \mid e = -1]$, i.e.

$$u(\overline{c}(1 - bL), -1) = u(\overline{c}, 0) = \phi(\overline{c})$$

and so

$$u(c, -1) = \phi(c/(1 - bL)) = \phi(cg(-1))$$

Therefore, $u(c, e) = \phi(cg(e))$ for all c and $e \in \{-1, 0\}$.

The converse of the proof is immediate, with $b = (1 - g(0)/g(-1))/L$.

Proof of Proposition 11   If perk consumption occurs, $P_1 = P - \frac{P}{S}\overline{L}$. For the manager not to take perks, we require

$$f + \nu P > \frac{f + \nu\left(P - \frac{P}{S}\overline{L} \right)}{1 - \Lambda},$$

and so $f\Lambda < \nu P \left(\frac{\overline{L}}{S} - \Lambda \right)$. Since $f \geq 0$, this cannot be satisfied if $S > \overline{L}/\Lambda$.

C   Theory Complements

C.1   Multiperiod Model

This Appendix underpins Section 2.4, which extends the pay-performance sensitivity results of Sections 2.1-2.3 to wealth-performance sensitivity in an intertemporal framework. We use the setup of Kreps-Porteus (1978), Epstein-Zin (1990) and Weil (1989), so that we have risk neutrality and smooth consumption over time.[28] Let the value function V_t denote the discounted utility of future consumption:

$$\ln V_t = (1 - \delta) \ln(c_t) + \delta \ln E_t[V_{t+1}] - \Lambda e_t \Delta t.$$

---

[28] As in the core model, risk neutrality significantly enhances tractability (and thus calibratability). Without smooth consumption, the model would be degenerate as the CEO consumes everything in a period in which he shirks.

For instance, if consumption and effort are deterministic, $\ln V_t = \sum_{s=0}^{\infty} \delta^s \left((1-\delta) \ln c_{t+s} - \Lambda e_{t+s} \right).$[29]

For simplicity, we assume $\delta = 1/(1+r_f)$, where r_f is the equilibrium riskless rate. Let W_t denote the CEO's wealth (financial wealth F_t plus the NPV of future pay). The optimal consumption policy is $c_t = r_f W_t / (1+r_f)$. The model is most suited for a continuous time setup, but for expositional reasons, we proceed in discrete time and take the continuous time limit where applicable.

The CEO has a fraction θ_t of his wealth in the firm. The firm's return is $r_{t+1} = r_f + Le_t + \eta_{t+1}$, where r is the risk-free rate and $e_t \in \{-1, 0\}$. Wealth evolves according to:

$$W_{t+1} = W_t \left(1 + r_f + \theta_t Le_t + \theta_t \eta_{t+1} \right) - c_{t+1}. \tag{47}$$

It is well-known that with a logarithmic utility function, the indirect utility of wealth is $\ln V_t = \ln W_t + k$, where k is a constant independent of wealth.

We now address the incentive compatibility condition. If the CEO shirks at time t, he increases his utility $\ln V_t$ by $\Lambda \Delta t$. On the other hand, his wealth at $t+1$ is lower by: $\Delta W_t = -W_r(t) L \Delta t$, where $W_r = \partial W / \partial r$. (In our example, $W_r = W\theta$.) Given that the utility is $\ln V_t = \ln W_t + k$, shirking increases utility $\ln V_t$ by:

$$\delta \ln V_t = \Lambda \Delta t + \ln(W_t + \Delta W_t) - \ln W_t = \Lambda \Delta t + \ln \left(1 - \frac{W_r(t) \cdot L \Delta t}{W_t} \right) = \Delta t \left(\Lambda - \frac{W_r(t) L}{W_t} \right) + o(\Delta t).$$

We take the continuous time limit, $\Delta t \to 0$. The agent does not shirk if and only if: $\Lambda - \frac{W_r(t)L}{W_t} \leq 0$, i.e.:

$$\frac{\partial W}{\partial r} \geq \frac{\Lambda}{L} W \tag{48}$$

As in Section 2, we select the contract that minimizes the risk in the CEO's pay. It is given by

$$\frac{\partial W}{\partial r} = \frac{\Lambda}{L} W.$$

Using Definition 2, the wealth-performance sensitivities in Proposition 6 can be easily derived.

We can further analyze CEO wealth. Assume pay grows at a rate g, so that $w_t = w_0 e^{gt}$, and the CEO exits the labor market with Poisson probability λ. Then, the NPV of future pay is: $\int_0^{\infty} e^{-r_f s} e^{-\lambda s} w_{t+s} ds = \int_0^{\infty} e^{-(r_f + \lambda)s} w_t e^{gs} ds = w_t / (r_f + \lambda - g)$ and total wealth (NPV of future wages, plus financial wealth) is $W_t = w_t / (r_f + \lambda - g) + F_t$.

---

[29] This is still a multiplicative model, like (1). The non-log analog would be:

$$V_t = \left[(1-\delta) c_t^{1-\sigma} + \delta \left(E_t [V_{t+1}] \right)^{1-\sigma} \right]^{1/(1-\sigma)} (1 - \Lambda e_t \Delta t)$$

as shirking for 1 period increases utility only by an amount proportional to $\Lambda \Delta t$.

Proposition 12 *(Pay for performance sensitivities in the intertemporal model, more explicit version). Let Λ denote the cost of effort, L the impact on firm value, g the expected growth rate of pay, λ the probability of the CEO exiting the labor market, F_t his financial wealth, and w_t his expected pay. Then the equilibrium wealth-performance sensitivities, defined in Definition 2, are:*

$$B^I = \frac{\Lambda}{L}\left(\frac{1}{r_f + \lambda - g} + \frac{F_t}{w_t}\right) \tag{49}$$

$$B^{II} = \frac{\Lambda}{L}\left(\frac{w_t}{r_f + \lambda - g} + F_t\right)\frac{1}{S_t} = B^I \frac{w_t}{S_t}$$

$$B^{III} = \frac{\Lambda}{L}\left(\frac{w_t}{r_f + \lambda - g} + F_t\right) = B^I w_t.$$

The scalings with firm size S and the size of the reference firm S_* are as in Proposition 5.

While equation (48) made predictions about the "stock" of incentives, we also wish to examine the flow of incentives, i.e. the optimal composition of the CEO's incremental compensation next period. Let W^Δ denote the increment in wealth brought by the new compensation. Assume no consumption for simplicity, and that currently $\frac{\partial W}{\partial r} \geq \frac{\Lambda}{L}W$ so that incentive compatibility is achieved. The CEO's new wealth is $W' = W + W^\Delta$. To maintain incentive compatibility, we require $\frac{\partial W^\Delta}{\partial r} \geq \frac{\Lambda}{L}W^\Delta$, and so $\frac{\partial W'}{\partial r} \geq \frac{\Lambda}{L}W'$. The least risky contract satisfying this condition is given by:

$$\frac{\partial W^\Delta}{\partial r} = \frac{\Lambda}{L}W^\Delta. \tag{50}$$

The one-period model of Section 2 predicted exactly (50). Hence, if one accepts the above selection criterion, then the predictions we obtain for the incentive mix in the flow of compensation are exactly the same as in the one-period model of Section 2, in particular Propositions 3, 4, and 5.

C.2 Perks in the Extended Model

We consider the wealth-performance sensitivity required for the CEO to sufficiently suffer from the 0.1% negative return to deter perk consumption. Let the perk of $\$\overline{L}$ be worth $\$\lambda\overline{L}$ to the CEO. λ parameterizes the inefficiency of perk consumption, where $0 < \lambda < 1$ so that perk consumption is inefficient.[30] We use the optimal incentive scheme of Proposition 7, which derives a constant portfolio share in the firm. Perk consumption increases the CEO's utility by $\lambda\overline{L}$ and reduces the stock return by $\overline{L}/S$, hence his wealth by $W\theta\overline{L}/S$. He therefore avoids the perk if and only if $\lambda\overline{L} - W\theta\overline{L}/S \leq 0$, i.e. $\theta \geq \lambda\frac{S}{W}$. With $\lambda = 1/2$, and $W = \$100$ million, we obtain a portfolio share $\theta \geq 50$. This implies that the CEO must invest 5,000% of his wealth

---

[30] λ is related to our earlier variable by $\lambda = \Lambda W/\overline{L}$.

in the firm, borrowing to reach that amount (this is possible in continuous time, while always maintaining positive wealth). This is clearly extreme, for any non-trivial level of risk aversion.

More precisely, we now use a total surplus perspective to explicitly account for the cost of risk-bearing and show that it substantially outweighs the benefit of perk prevention. Perk consumption reduces firm value by $\overline{L}$ per unit of time and increases the CEO's utility by $\lambda \overline{L}$, so the net loss is $(1-\lambda)\overline{L}$. If firm chooses to deter perks with incentives (rather than direct control), it needs to implement a portfolio share $\theta = \lambda \frac{S}{W}$, so the loss to total surplus calculated from (27), is (per unit of time): $W\Gamma\theta^2\frac{\sigma^2}{2} = \lambda^2\frac{S^2}{W}\Gamma\frac{\sigma^2}{2}$. Total surplus rises with perk prevention if and only if $(1-\lambda)L \geq \lambda^2\frac{S^2}{W}\frac{\Gamma\sigma^2}{2}$, i.e.

$$\frac{\overline{L}W}{S^2} \geq \frac{\lambda^2}{1-\lambda}\frac{\Gamma\sigma^2}{2}.$$

With $\Gamma = 1$, $\sigma^2 = 0.04$ (an annual volatility of 20%) and $\lambda = 1/2$, the right-hand side is equal to 1%. The left-hand size is (\$10 million)(\$100 million)/(\$10^4 million)^2 = 10^{-5}. Hence, the losses from risk-bearing are several orders of magnitude higher than the gains from perk prevention. As with the previous subsection, this implies that perks are best controlled through active corporate governance. The exception is for very small firms, where W is of a similar magnitude to S.

We summarize this result in the next Proposition:

Proposition 13 *(Perk prevention with general incentive contracts). Perks can be deterred with general incentive contracts if the CEO receives a share:*

$$\theta \geq \lambda\frac{S}{W}. \tag{51}$$

It is inefficient to deter perk consumption if and only if:

$$\frac{\overline{L}W}{S^2} \leq \frac{\lambda^2}{1-\lambda}\frac{\Gamma\sigma^2}{2}. \tag{52}$$

References

[1] Aghion, Philippe and Patrick Bolton (1992): "An Incomplete Contracts Approach to Financial Contracting." *Review of Economic Studies* 59, 473-94

[2] Almazan, Andres and Javier Suarez (2003): "Entrenchment and Severance Pay in Optimal Governance Structures." *Journal of Finance* 58, 519-548

[3] Armstrong, Chris, David F. Larcker, Che-Lin Su (2007): "Stock Options and Chief Executive Officer Compensation." Working Paper, Stanford University

[4] Baker, George, Michael Jensen and Kevin J. Murphy (1988): "Compensation and Incentives: Practice vs. Theory." *Journal of Finance* 43, 593-616

[5] Baker, George and Brian Hall (2004): "CEO Incentives and Firm Size." *Journal of Labor Economics* 22, 767-98

[6] Baranchuk, Nina, Glenn MacDonald and Jun Yang (2007): "The Economics of Super Managers." Working Paper, University of Texas at Dallas.

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

[8] Bebchuk, Lucian, and Yaniv Grinstein (2005): "The Growth of Executive Pay." *Oxford Review of Economic Policy* 21, 283-303

[9] Becker, Bo (2006): "Wealth and Executive Compensation." *Journal of Finance* 61, 379-397

[10] Bertrand, Marianne and Sendhil Mullainathan (2001): "Are CEOs Rewarded for Luck? The Ones Without Principals Are." *Quarterly Journal of Economics* 116, 901–932

[11] Core, John and Wayne Guay (1999): "The Use of Equity Grants to Manage Optimal Equity Incentive Levels." *Journal of Accounting and Economics* 28, 151-184

[12] Core, John and Wayne Guay (2002a): "Estimating the Value of Employee Stock Option Portfolios and Their Sensitivities to Price and Volatility". *Journal of Accounting Research* 40, 613-630

[13] Core, John, Wayne Guay, and Robert Verrecchia (2003): "Price Versus Non-Price Performance Measures in Optimal CEO Compensation Contracts." *The Accounting Review* 78, 957-981

[14] Demsetz, Harold and Kenneth Lehn (1985): "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93, 1155-1177

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

[16] Edmans, Alex (2007): "Inside Debt". Working Paper, University of Pennsylvania

[17] Epstein, Larry and Stanley Zin (1989): "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57, 937-968

[18] Falato, Antonio, and Dalida Kadyrzhanova (2007) "Optimal CEO Incentives and Industry Dynamics." Working Paper, Federal Reserve Board

[19] Frydman, Carola and Raven Saks (2007): "Historical Trends in Executive Compensation, 1936-2003." Working Paper, Massachusetts Institute of Technology

[20] Gabaix, Xavier and Augustin Landier (2008): "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, forthcoming.

[21] Garicano, Luis and Thomas Hubbard (2005): "Learning About the Nature of Production from Equilibrium Assignment Patterns." Working Paper, University of Chicago

[22] Gayle, George-Levi and Robert Miller (2007): "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" Working Paper, Carnegie-Mellon University

[23] 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

[24] Gompers, Paul, Joy Ishii and Andrew Metrick (2003): "Corporate Governance and Equity Prices." *Quarterly Journal of Economics* 118, 107-155

[25] Grossman, Sanford J. and Oliver D. Hart (1983): "An Analysis of the Principal-Agent problem," *Econometrica* 51, 7-45

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

[27] Haubrich, Joseph G. (1994): "Risk Aversion, Performance Pay, and the Principal-Agent Problem." *Journal of Political Economy* 102, 258-276.

[28] He, Zhiguo (forthcoming) "Optimal Executive Compensation When Cash Flows Follow Geometric Brownian Motion," *Review of Economic Studies* .

[29] Hermalin, Benjamin and Michael Weisbach (1998): "Endogenously Chosen Boards of Directors and Their Monitoring of Management." *American Economic Review* 88, 96-118

[30] Himmelberg, Charles and Glenn Hubbard (2000): "Incentive Pay and the Market for CEOs: An Analysis of Pay-for-Performance Sensitivity." Working Paper, Columbia University

[31] Holmstrom, Bengt (1979): "Moral Hazard and Observability." *Bell Journal of Economics* 10, 74-91

[32] Holmstrom, Bengt (1992): "Comments," in Lars Werin and Hans Wijkander, eds., *Contract Economics.* Blackwell, Cambridge MA

[33] Holmstrom, Bengt and Steven Kaplan (2001): "Corporate Governance and Takeovers in the US: Making Sense of the '80s and '90s." *Journal of Economic Perspectives* 15, 121-144

[34] Holmstrom, Bengt and Paul Milgrom (1987): "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55, 308-328

[35] Inderst, Roman and Holger Mueller (2007): "CEO Compensation and Private Information: An Optimal Contracting Perspective." Working Paper, New York University

[36] Jensen, Michael and William Meckling (1976): "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure." *Journal of Financial Economics* 3, 305-360

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

[38] Jensen, Michael and Kevin J. Murphy (2004): "Remuneration: Where We've Been, How We Got To Here, What Are The Problems, and How To Fix Them." Working Paper, Harvard Business School

[39] Kreps, David and Evan Porteus (1978): "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46, 185-200.

[40] Kuhnen, Camelia and Jeffrey Zwiebel (2007): "Executive Pay, Hidden Compensation and Managerial Entrenchment." Working Paper, Northwestern University

[41] Manso, Gustavo (2006): "Motivating Innovation." Working Paper, Massachusetts Institute of Technology

[42] Mehra, Rajnish and Edward Prescott (1985): "The Equity Premium: A Puzzle?" *Journal of Monetary Economics* 15, 145-161

[43] Morck, Randall, Andrei Shleifer and Robert Vishny (1990): "Do Managerial Objectives Drive Bad Acquisitions?" *Rand Journal of Economics* 24, 357-379

[44] Murphy, Kevin J. (1985): "Corporate Performance and Managerial Remuneration: An Empirical Analysis." *Journal of Accounting and Economics*, 7, 11–42

[45] Murphy, Kevin J. (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

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

[47] Rajan, Raghuram and Julie Wulf (2006): "Are Perks Purely Managerial Excess?" *Journal of Financial Economics* 79, 1-33

[48] Rosen, Sherwin (1992): "Contracts and the Market for Executives," in Lars Werin and Hans Wijkander, eds., *Contract Economics*. Cambridge, MA and Oxford: Blackwell, 181–211

[49] Sannikov, Yuliy (2006): "A Continuous-Time Version of the Principal-Agent Problem." Working Paper, New York University

[50] Sattinger, Michael (1993): "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature* 31, 831–880

[51] Schaefer, Scott (1998): "The Dependence of Pay-Performance Sensitivity on the Size of the Firm." *Review of Economics and Statistics*, 80, 436-443

[52] Tervio, Marko (2007): "The Difference that CEOs Make: An Assignment Model Approach." Working Paper, University of California, Berkeley

[53] Weil, Philippe (1990): "Nonexpected Utility in Macroeconomics." *Quarterly Journal of Economics* 105, 29-42

[54] Yermack, David (1995): "Do Corporations Award CEO Stock Options Effectively?" *Journal of Financial Economics* 39, 237-269

[55] Yermack, David (2006): "Flights of Fancy: Corporate Jets, CEO Perquisites, and Inferior Shareholder Returns". *Journal of Financial Economics* 80, 211-242

Table 1: Comparing Different Measures of Pay-Performance Sensitivity.

	b^{I}	b^{II}	b^{III}
Measures	$\dfrac{\Delta \ln c}{\Delta \ln S}$	$\dfrac{\Delta c}{\Delta S}$	$\dfrac{\Delta c}{\Delta \ln S}$
Real variables	$\frac{\text{\$shares}}{\text{total pay}}$	% shares	$ shares
WPS analog	$\dfrac{\Delta \$W}{\Delta \ln S}\dfrac{1}{w}$	$\dfrac{\Delta \$W}{\Delta \$S}$	$\dfrac{\Delta \$W}{\Delta \ln S}$
Used by	Murphy (1985)	Demsetz-Lehn (1985)	Holmstrom (1992)
	Rosen (1992)	Jensen-Murphy (1990)	Hall-Liebman (1998)
	Boschen-Smith (1995)	Yermack (1995)	
		Schaefer (1998)	
This paper	$\dfrac{\Lambda}{L}$	$\dfrac{\Lambda}{L}\dfrac{w}{S}$	$\dfrac{\Lambda}{L}w$
Scaling with S	$b^{I} \propto S^{0}$	$b^{II} \propto S^{\rho-1}$	$b^{III} \propto S^{\rho}$
	$b^{I} \propto S^{0}$	$b^{II} \propto S^{-2/3}$	$b^{III} \propto S^{1/3}$
Scaling with $S(n_{*})$	$b^{I} \propto S^{0}S(n_{*})^{0}$	$b^{II} \propto S^{-(1-\rho)}S(n_{*})^{\gamma-\rho}$	$b^{III} \propto S^{\rho}S(n_{*})^{\gamma-\rho}$
	$b^{I} \propto S^{0}S(n_{*})^{0}$	$b^{II} \propto S^{-2/3}S(n_{*})^{2/3}$	$b^{III} \propto S^{1/3}S(n_{*})^{2/3}$

Explanation: This Table shows the 3 different measures of pay-performance sensitivity (WPS denotes wealth-performance sensitivity). c is the realized compensation, w is the expected compensation, S is the market value of the firm, Λ is the disutility from effort, L is the value lost from shirking, and W is the wealth. We suppress the dependence on firm n for brevity. ρ is the cross-sectional elasticity of expected pay to firm size ($w \propto S^{\rho}$) and empirically is around $\rho = 1/3$. The predictions in this table are from Propositions 3, 4 and 5. The symbol "$\propto$" denotes "is proportional to". For instance, $b^{II} \propto S^{-2/3}$ means that we predict that b^{II} declines with size S, with an elasticity of -2/3.

Table 2: Elasticities of Pay-Performance Sensitivity with Firm Size.

	$\ln(B^{I})$	$\ln(B^{II})$	$\ln(B^{III})$
ln(Aggregate Value)	0.0648	-0.5778	0.4222
	(0.0671)	(0.0526)	(0.0526)
Year Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Firm Fixed Effects	No	No	No
Observations	5,973	5,973	5,973
Adj. R-squared	0.1718	0.3453	0.3618

Explanation: We merge Compustat with ExecuComp (1992-2005) and select the 500 largest firms each year by aggregate value (debt plus equity). We use the Core and Guay (2002a) methodology to estimate the delta of the CEO's option holdings. B^{I}, B^{II} and B^{III} are estimated using equations (22)-(24). See Appendix A for full details. The industries are the Fama-French (1997) 48 sectors. Standard errors, displayed in parentheses, are clustered at the firm level. Based on the calibration of Gabaix and Landier (2008), the model predicts an elasticity of $\rho = 0$ for B^{I}, $\rho = -2/3$ for B^{II}, and $\rho = 1/3$ for B^{III}.

Table 3: The Positive Relation between Compensation Volatility and Firm Volatility.

	Ex ante measure of volatility			Ex post measure of volatility		
	$\ln(B^I \sigma_r)$	$\ln(B^{II} \sigma_r)$	$\ln(B^{III} \sigma_r)$	$\ln\left(\frac{|W_{t+1}-W_t|}{w_t}\right)$	$\ln\left(\frac{|W_{t+1}-W_t|}{S_t}\right)$	$\ln|W_{t+1}-W_t|$
ln(return vol)	1.0882	1.3327	1.3327	0.6435	0.9659	0.9714
	(0.1322)	(0.1199)	(0.1199)	(0.1816)	(0.1550)	(0.1584)
ln(firm size)	0.0705	-0.5564	0.4436	0.0346	-0.5679	0.4045
	(0.0686)	(0.0539)	(0.0539)	(0.0691)	(0.0552)	(0.0560)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Firm FE	No	No	No	No	No	No
Observations	5,973	5,973	5,973	4,035	4,035	4,035
Adj. R-squared	0.2586	0.4478	0.4508	0.1421	0.2916	0.2790

Explanation: We merge Compustat with ExecuComp (1992-2005) and select the 500 largest firms each year by aggregate value (debt plus equity). We use the Core and Guay (2002a) methodology to estimate the delta of the CEO's option holdings. B^I, B^{II} and B^{III} are estimated using equations (22)-(24). See Appendix A for full details. The industries are the Fama-French (1997) 48 sectors. Standard errors, displayed in parentheses, are clustered at the firm level. The theory predicts a positive coefficient between wealth volatility and stock-return volatility, contrary to additive models with unbounded effort.