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

August 21, 2007

Abstract

This paper presents a unified framework for understanding the determinants of both CEO incentives and total pay levels in competitive market equilibrium. It embeds a modified principal-agent problem into a talent assignment model to endogenize both elements of compensation. The model's closed form solutions yield testable predictions for how incentives should vary across firms under optimal contracting. In particular, our calibrations show that the negative relationship between the CEO's effective equity stake and firm size is quantitatively consistent with efficiency and need not reflect rent extraction. Our model and data both also imply that the dollar change in wealth for a percentage change in firm value, scaled by annual pay, is independent of firm size. This may render it an attractive incentive measure as it is comparable between firms and over time. The theory also predicts a positive relationship between pay volatility and firm volatility, and that risk and effort affect total pay along the cross-section but not in the aggregate. Finally, we demonstrate that incentive compensation is effective at solving large agency problems, such as selecting corporate strategy, but smaller issues such as perk consumption are best addressed through direct monitoring. (JEL: D2, D3, G34, J3)

*aedmans@wharton.upenn.edu, xgabaix@stern.nyu.edu, alandier@stern.nyu.edu.

For helpful comments, we thank John Core, Carola Frydman, Bob Gibbons, Ben Hermalin (a discussant), Dirk Jenter, Holger Mueller, Michael Roberts, Yuliy Sannikov, Andrei Shleifer, Jeremy Stein, Youngsuk Yook and seminar participants at MIT, NYU and the NBER Summer Institute. XG thanks the NSF (DRU-0527518) for financial support.

This paper presents a unified framework for understanding the determinants of both the level and sensitivity of CEO pay in neoclassical market equilibrium. In our model, both elements of compensation are simultaneously governed by the market for scarce talent and the nature of the agency conflict. Holding total pay constant, effort considerations determine its division into fixed and performance-sensitive components. To endogenize the level of total pay, and thus fully solve for the absolute level of incentive compensation, we embed this result into a general equilibrium model of the competitive assignment of CEO talent. As in Gabaix and Landier (2008), 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. Dollar incentive compensation therefore also varies with size. We further extend the competitive assignment model to incorporate risk aversion and allow for general contracts, deriving further implications on the effect of risk on pay and the effectiveness of compensation in addressing agency problems.

The model has two key contributions over and above existing theories. First, while many incentive models are partial equilibrium, taking the level of pay as given, we endogenize it in market equilibrium to produce a single, parsimonious model of both incentives and total pay. The model therefore combines many issues related to executive compensation in a single framework, demonstrating how incentives and salary should optimally vary across companies, between countries, and over time according to managerial talent, firm size, volatility and the cost of effort. Second, our model is particularly tractable and yields closed-form solutions. These features remain when allowing for general incentive contracts and general risk-averse utility functions. Indeed, the full market equilibrium can be summarized in just three simple equations. Not only may this make the model an attractive benchmark on which future theories can build, but it also leads to clear, quantitative empirical implications and thus readily lends the model to empirical analysis.

We explore three such implications. The first is the relationship between firm size and wealth-performance sensitivity. This issue is important for at least two reasons. It has been widely documented that the CEO's "effective equity stake" or "dollar-dollar incentives" (the dollar change in wealth for a dollar change in firm value) are 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. In our theory, effort has a multiplicative effect on firm value, and so the dollar gains from working are proportional to size. The CEO's utility gain from shirking (in dollar terms) rises with wealth, but wealth only has a 1/3 elasticity with size. Therefore, dollar-dollar incentives should have a size elasticity of -2/3, which is very close to our empirical estimate of -0.58. Therefore, the observed negative relationship is exactly what a frictionless model would predict – a smaller effective equity share is sufficient to induce effort in large companies. Note that 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.

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, as it is comparable across firms and over time.

Second, we examine the model's implications for the effect of firm risk on total pay, the level of incentives, and pay volatility. Traditional theories have an unbounded level of effort and so optimal incentives are a trade-off between the gains from working and the cost of risk-bearing. Firm risk therefore reduces both the level of incentives and pay volatility. Our model features a maximum level of effort, which the firm always wishes to implement as the gains from effort are proportional to firm size, but the cost imposed on the CEO is proportional to his wage, which is substantially smaller. Incentives are set to induce maximum effort regardless of risk, and so are independent of volatility. Since pay volatility equals the product of incentives and firm risk, we predict a positive link between pay volatility and risk, contrary to existing models but supported by our data.

For total pay, the theory predicts that variations in volatility generate cross-sectional salary differences as riskier firms have to pay a compensating differential. We confirm this empirically. However, market-wide increases in risk have negligible impact on the pay of the most talented CEOs. Since the pay of top CEOs is only driven by firm size and the scarcity of CEO talent, it is not compensation for aggregate-level risk. The effects of disutility of effort are very similar – it explains pay differences along the cross-section, but has no aggregate impact.

The third application of the model is to assess whether observed levels of incentive compensation are effective in solving agency problems. Jensen and Murphy (1990) find that CEO wealth falls by only $3.25 for every $1,000 loss in shareholder value. As this figure appears low, it is frequently interpreted as evidence that current practices are inadequate to induce shareholder value maximization (see however Hall and Liebman (1998)). Since this issue concerns magnitudes, not directions, a calibratable model is particularly suited to shed light upon the debate. We find that observed incentives are able to deter suboptimal actions (such as shirking, pursuit of pet projects, or empire-building acquisitions) if such behavior increases the CEO's utility by a monetary equivalent no greater than 0.9 times his annual wage. Since it appears plausible that the private benefits from most potential value-destructive actions fall below this upper bound, incentives are able to solve the majority of agency problems. Apparently small incentives can have substantial power because 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 very high and so the manager only needs a small equity stake to achieve incentive compatibility.

While the above calibration focuses on the potency of current levels of incentives, a related contribution is to analyze the effectiveness of incentives in general (for any reasonable levels) at addressing agency problems. The seminal model of Jensen and Meckling (1976) implies that all agency issues can and should be solved by incentives, but we show that there are certain problems for which compensation is ineffective. First, some actions may yield the CEO substantial private benefits, which may exceed the loss in wealth implied from any plausible level of incentives. One example is managerial entrenchment – by failing to (optimally) resign voluntarily, the CEO may enjoy his salary and private benefits of control for many future years. Second, some actions may have too small an effect on the firm's stock returns for the CEO's equity holdings to be sensitive. The core model considers actions which have a multiplicative effect on firm value (such as changes in strategy) and thus affect stock returns, regardless of firm size. However, certain actions such as perk consumption (e.g. the purchase of a corporate jet) reduce firm value by a fixed dollar amount, independent of size, and thus have a very small effect on the returns of a large company. The manager's equity stake is thus insufficient to deter perks. When we allow for general contracts, perks can be deterred by using extremely sensitive instruments, but these impose such a large risk-bearing cost on the manager that total surplus falls. Hence, in our model, perk prevention has no explanatory power for incentive compensation, and can only achieved through active corporate governance, e.g. direct rules imposed on the CEO. Incentive compensation is effective at solving large agency problems with a significant impact on returns, but smaller issues such as perk consumption are best addressed through direct monitoring.

While individual predictions may be achievable from alternative models, to our knowledge the combination of the above implications, plus the relationships between total pay and firm size stated in Gabaix and Landier (2008), are unique to our unifying framework. Uniting all of these predictions in a single parsimonious model is not the only advantage of endogenizing both total pay and incentives together. Our market 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. Even when risk aversion is introduced, incentive considerations in the aggregate change the sensitivity of pay to performance, but not expected pay. Conversely, talent determines the level of pay but not its incentive component.

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, a first-best benchmark is necessary to evaluate whether they are now "high enough."

The most closely related theory papers are calibrations of the CEO incentive problem. While the main 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 effectiveness of incentives at solving different types of agency problems.

Our paper differs from the above papers owing to its contrasting objectives (principally, the effect of size on incentives) and its modeling approach (general equilibrium incorporating both pay and incentives). 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 total pay, and their scaling with firm size and volatility.

In contemporaneous work, Baranchuk, Macdonald and Yang (2007) and Falato 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 1 we model equilibrium compensation for a risk-neutral CEO, generating predictions for the effect of size on incentives. Section 2 studies the optimal contract for a risk-averse CEO and explores the effect of risk and cost of effort on pay. Section 3 presents empirical evidence quantitatively consistent with the model's main predictions for the scaling of incentives with firm size. Section 4 considers further implications of the model and Section 5 concludes.

1 The Basic Model

We start in Section 1.1 by deriving the optimal division of CEO compensation into stock and cash salary, in a partial equilibrium analysis that takes total compensation as given. In Section 1.2 we embed this analysis into a general equilibrium where total pay is endogenously determined, and present the implications for pay-performance sensitivity in Section 1.3. Section 1.4 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, we maximize tractability by building 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. In addition, risk neutrality 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 match the data. Section 2 will later show that our predictions are robust to relaxing these assumptions, and analyze the effect of risk on compensation.

1.1 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 $e \in \{\underline{e}, \overline{e}\}$ denotes CEO effort. We normalize $\overline{e} = 0$, $g\left(\overline{e}\right) = 1$ and set $g\left(\underline{e}\right) = 1/\left(1 + \Lambda \underline{e}\right)$, where $\Lambda \in [0, 1)$. Shirking reduces firm value by a fraction $\underline{e}$ and increases the CEO's utility by (approximately) a fraction $\Lambda \underline{e}$. Λ parameterizes the effort cost required to increase firm value by a given amount, which we will refer to this as the "unit cost 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 1.2.

Equation (1) is a multiplicative functional form, generalized in Section 2 to other forms such as $E\left[u\left(cg\left(e\right)\right)\right]$. We use this specification as it it seems highly plausible that the utility gains from shirking are increasing in the CEO's wage. For example, shirking allows the CEO to enjoy consuming goods and services that he can purchase with his salary, and so leisure and consumption are complementary goods. Multiplicative preferences mean that the share of total pay allocated to consumption and leisure is independent of the wage – changes in salary do not affect the composition of the "bundle" of consumption and leisure purchased by the CEO, only the overall size of the bundle. In addition to a positive consumption-leisure relationship being psychologically appealing, it also has empirically consistent implications for the scaling of labor supply with the wage, since it implies labor supply does not have diverging trends over time.[1] This empirical consistency explains its common use in macroeconomics, a field in which models are frequently calibrated to the data. By contrast, the additive functional forms commonly used in qualitative corporate finance models (such as $E\left[c^{\alpha}\right]-g\left(e\right)$) are both arguably less plausible (implying that the benefits from shirking are independent of the wage) and have empirically inconsistent implications, such as predicting that leisure falls to zero as the wage rises over time. In Section 4.3 we detail further counterfactual predictions of additive preferences.

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

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

where η is stochastic noise with mean 0. Low effort ($e=\underline{e}$) reduces firm value by a fraction $\underline{e}$.

We assume that $S > w\Lambda$, 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 action that increases firm value but involves a non-pecuniary cost to the manager. In the literal interpretation, $e=0$ represents high effort and $e=\underline{e}$ is shirking. A second interpretation is the choice of an investment project, strategy or acquisition target, where $e=0$ is the first best project and $e=\underline{e}$ yields the CEO private benefits, such as an empire-building expansion. The effects of effort or project choice plausibly have a proportional effect on firm value, explaining the formulation in equation (2). However, certain actions have a fixed dollar effect independent of firm size, such as perk consumption or managerial rent extraction through stealing. We consider such additive actions

[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.

[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 $S(1+\overline{e})-w$, leading to total surplus of S. If the manager shirks, he is paid $w(1+\Lambda\underline{e})$ (to keep his utility at w). Firm value (net of wages) is $S(1+\underline{e})-w(1+\Lambda\underline{e})$ and total surplus is $S(1+\underline{e})-w\Lambda\underline{e}$. Hence total surplus is higher if the manager works if and only if $S>w\Lambda$.

in Section 4.1.

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* $\Lambda < 1$ *(the cost of effort is not too strong). The optimal contract pays a fraction* Λ *of the wage in shares, and the rest in cash. Namely, it comprises a fixed base salary,* f^**, and* $\nu^* P$ *worth of shares, with:*

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

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

where Λ *is the unit cost of effort. The manager's realized compensation is:*

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

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.

1.2 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)

[4]Section 2 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.

("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:[5]

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

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 1.1 to determine the form of compensation. We allow Λ to differ across firms, and so index it Λ_n. We need not make any assumptions on how Λ_n varies with n: as long as $\Lambda_n < 1$ 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* $\forall\ n,\ \Lambda_n < 1$ *(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:*

$$\begin{aligned} \nu_n^* P_n &= w(n)\Lambda_n, \\ f_n^* &= w(n)(1-\Lambda_n), \end{aligned}$$

where Λ_n *is the manager's disutility of effort. The manager's realized compensation is:*

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

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

[5] Throught this paper, we consider the domain of very large firms, i.e. take the limit $n/N \to 0$, where N is the total mass of firms.

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 currently risk-neutral. 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.[6] As long as pay would have been even higher had the manager delivered stronger performance, it can be consistent with optimal contracting.

1.3 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 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 $\Lambda_n = \Lambda$ across all firms.[7]

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

[6] 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.

[7] 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).

given by:

$$b^I = \Lambda \tag{12}$$
$$b^{II} = \Lambda \frac{w}{S} \tag{13}$$
$$b^{III} = \Lambda w, \tag{14}$$

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} denotes 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:*

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

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,

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

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:*

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

where γ *is the elasticity of CEO impact in GL (equation (38)). In particular, in the calibration* $\rho = 1/3, \gamma = 1$, *used in GL,*

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

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 independent 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 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 scales with $S^{1/3}$ in market equilibrium, 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 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 1.1 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.

While this paper models incentive pay as the solution to an effort problem, incentives can be used for alternative purposes such as screening out low-ability CEOs (Lazear (1995), Holmstrom (1999)) or indexing the CEO wage to market conditions (Oyer 2004). In future work, it might be interesting to analyze variants of the model that incorporate other reasons for incentive pay and explore the resulting empirical implications. By seeing which model's predictions most closely match the data, we may understand better the main motivations for incentive pay in practice: solving agency problems, screening, or alternative theories.

1.4 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 B 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:*

$$B^{I} = \frac{\partial W}{\partial r}\frac{1}{w} = \frac{\Delta \$Wealth}{\Delta \ln Firm\ Value}\frac{1}{\$Wage} \tag{16}$$

$$B^{II} = \frac{\partial W}{\partial r}\frac{1}{S} = \frac{\Delta \$Wealth}{\Delta \$Firm\ Value} \tag{17}$$

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

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 *plus* the median return applied to the CEO's existing portfolio of shares and options.[8]

[8] 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.

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:*

$$B^{I} = \Lambda \frac{W}{w} \tag{19}$$

$$B^{II} = \Lambda \frac{W}{S} \tag{20}$$

$$B^{III} = \Lambda W. \tag{21}$$

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. Since shirking and consuming are complementary goods, higher wealth raises current consumption and thus the utility gains from shirking. Pay-performance sensitivity must therefore rise to continue to induce effort.

The numerical scalings for pay-performance sensitivity in equation (15) were obtained using the well-documented 1/3 elasticity of the wage with size. Using the data from Section 3, in unreported results we confirm that this elasticity holds for the relationship between wealth and size: we find a coefficient of 0.40 with a standard error of 0.05. By contrast, W/w has an elasticity of 0.04, less than its standard deviation. Note that we only have data on the CEO's financial wealth in his own firm (plus accumulated annual flow compensation), and so our results assume the proportion of own-firm financial wealth to total wealth is constant across firm size.

2 Extended Model with Risk Aversion and General Contracts

The previous section 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 2.1 considers the extended model in partial equilibrium and in Section 2.2 we embed it in market equilibrium.

2.1 Partial Equilibrium: A "Detail-Independent" Optimal Contract

Let $e \in [\underline{e}, \overline{e}]$ denote the CEO's effort. The end-of period firm return on assets, $R = P_1/P_0$, is:

$$R = (1+\eta) L(e) \tag{22}$$

where L is continuously differentiable, positive, increasing, and $\ln L$ is weakly concave. The maximum action is normalized to $L(\overline{e}) = 0$. η is a random disturbance outside the CEO's control. $\ln(1+\eta)$ has a bounded support. We assume that CEO sees the realization of η *before* choosing effort e, an assumption that will substantially simplifies the analysis.

The CEO's utility function is

$$u(cg(e)) \tag{23}$$

where c is terminal consumption, $g(e)$ captures the disutility of effort and is decreasing and positive, and $\ln g$ is concave. u has domain $\mathbb{R}_+^*$, is increasing and weakly concave, and $\lim_{c\to+\infty} u(c) = +\infty$. The CEO's reservation utility is $\underline{u}$].

The utility function (23) preserves and generalizes (1) in a number of ways. First, the utility function u can be a general concave function. Second, effort and consumption continue to affect each other multiplicatively rather than additively. Third, effort is no longer a binary variable.

We consider the case where the highest level of effort, $e = \overline{e}$, maximizes total surplus.[9] As before, this is optimal under weak assumptions, because the firm (and thus the benefit from effort) is very large compared to the CEO (and thus the cost of effort). The cost of effort now comprises both the direct disutility and the inefficient risk sharing that results from incentivizing the manager to exert effort.

At the maximum effort level, if the CEO increases firm returns by 1%, he decreases his utility (in consumption equivalent units) by Λ%, where:

$$\Lambda \equiv -\left(\ln g(\overline{e})\right)' / \left(\ln L(\overline{e})\right)' \tag{24}$$

As in Section 1, Λ represents the "cost of effort": the marginal rate of substitution between firm value and CEO utility

The CEO has a reservation utility $u(\underline{w})$ given by the competitive market, and we seek the optimal (unrestricted) contract, a function $c(R)$ of the realized return that implements $e = \overline{e}$, satisfies the participation constraint $U \geq \underline{u}$, and has the minimum cost $E[c]$ to the firm.[10] We

[9]Lemma 1 in Appendix A shows that this the case if the firm is sufficiently large.

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

can also allow compensation to depend on messages sent by the CEO to the firm, but as shown in the proof, they have no effect. The optimal contract is derived in Appendix A and stated below. [11]

Theorem 1 *(Optimal unrestricted contract, with a risk-averse CEO). The unrestricted optimal contract pays the CEO an amount W_1 defined by:*

$$W_1 = W_0 R^{\Lambda} \tag{25}$$

where W_0 is a constant than ensures that the participation constraint binds ($E\left[u\left(W_0 R^{\Lambda}\right)\right] = \underline{u}$) and R is the gross firm return at the end of the period. The functional form R^{Λ} is independent of the utility function u and the distribution of the noise η.

The contract in equation (25) has a simple practical implementation, in the case where firm returns follow a continuous-time diffusion between period 0 and 1. For simplicity of exposition, we normalize the interest rates and risk premia to 0. At time 0, the CEO is given a portfolio of value $E[W_1]$, of which a fraction Λ is invested in the stock and the remainder in cash. This portfolio is continuously rebalanced between periods 0 and 1, so that the fraction in the stock remains constant at Λ. The CEO's final wealth therefore becomes (25).[12]

Theorem 1 yields a particularly simple optimal contract. We describe it as "detail-independent" as its functional form does not depend on the distribution of the noise, nor on the CEO's utility function – these only affect the specific value of W_0. In particular, the shape of the optimal contract R^{Λ} depends only the cost of effort Λ, but not on risk aversion. This simple form contrasts with the great complexity of traditional contracts under risk aversion (e.g. Grossman and Hart (1983)).

The link with the optimal contract in Section 1 is as follows. Equation (25) can be rewritten $\ln W_1/W_0 = \Lambda \ln P_1/P_0$, so that $b^I = E\left[\partial \ln W/\partial r\right] = \Lambda$. Changes in log CEO wealth must be proportional to changes in log firm value, with a constant of proportionality of Λ. Therefore, final compensation is proportional to the stock price to the power Λ.

We conclude this subsection with some remarks on our model setup. Our framework makes three small departures from conventional models. First, we postulate multiplicative production and utility functions, which lead to scale-independent contracts. Second, the firm always wishes to implement maximum effort, since the benefits of effort outweigh the costs, which removes the need to analyze small trade-offs. Third, the CEO observes the realization of the noise before taking his action, which we believe to be plausible. The combination of these three

[11] If the CEO has any initial wealth, the contract is still given by (25), with a fraction Λ of both existing and new wealth being continuously invested in the stock.

[12] The proof is thus. The firm evolves as $dP_t/P_t = \sigma dz_t$ The CEO wealth V_t starts at $V_0 = E\left[W_1\right]$, and for $t \in [0,1]$, evolves at $dV_t/V_t = \sigma\Lambda dz_t$, so that $d\ln V_t = \sigma\Lambda dz_t - \sigma^2\Lambda^2 dt/2$, and the final value of the portfolio is $V_1 = E[W_1]\exp\left(\Lambda\sigma z_1 - \sigma^2\Lambda^2/2\right) = R^{\Lambda} W_0$.

departures leads to the particularly simple form of our optimal contract. Note that only the third assumption was deliberately made to maximize tractability; the first two were made as we believe they correspond most closely to the economics of the situation. The third assumption leads to tractability since, if the CEO observes η before choosing his effort level, the realization of η is immaterial for his decision problem. (This is shown most clearly in the first few lines of the proof). Therefore, the form of the contract is independent of the noise distribution. We suspect that, if we changed the third assumption, the qualitative features of the contract would be little affected; however, the solution would be substantially more complicated.[13]

We also note that, even though this is a hidden information model (the CEO learn the noise η before taking the action), there is no need for the CEO to send messages to the firm, and there is no need for "menus of contracts," as shown in detail in the proof. Intuitively, the reason is that the firm wishes to implement maximal effort in all cases. Hence, on the equilibrium path, there is a one to one correspondence between the firm's return and the noise, which makes messages redundant.

The simplicity of the contract in Theorem 1 allows it to be easily embedded in market equilibrium, a task to which we now turn.

2.2 Market Equilibrium

We now derive the market equilibrium with risk averse CEOs, using the optimal contract of the previous section. To obtain specific quantitative results, we specialize the utility function to $u(c) = c^{1-\Gamma}/(1-\Gamma)$ for $\Gamma \geq 0$, $\Gamma \neq 1$, and $u(c) = \ln c$ for $\Gamma = 1$. We take the return to be $R = \exp(e - \overline{e} + \sigma\varepsilon - \sigma^2/2)$, where ε is a standard Gaussian variable, so that $L(e) = \exp(e - \overline{e})$.[14] We normalize the risk premium and interest rate to 0, and take $g(e) = \exp(-\Lambda e)$, which is consistent with (24).[15] We allow for heterogeneity in the firm's cost of effort, scope of effort and volatility.[16] The CEO working for firm n receives an expected wage w_n. His utility is given by:

$$U = u(w_n \exp(-\chi_n)),$$

[13] In parallel work, we extend the model to continous time dynamic contracts, using the techniques of Sannikov (2006). The optimal contract remains equation (25). Hence the result of Theorem 1 is confirmed, which shows that its economics do not depend essentially on the details of the temporal resolution of uncertainty. It also may explain the superficial similarity between our contract, where log pay is affine in log performance, and that of Holmstrom and Milgrom (1987), where pay is affine in performance, even though our setup cannot be mapped into that of Holmstrom and Milgrom.

[14] Formally speaking, this Gaussian distribution of ε is unbounded, contradicting an assuption made in section 2.1. One can approach that condition arbitrarily closely, by truncating the distribution of ε to $-[A, A]$, for some very large, but finite, upper bound.

[15] If the firm's earnings are a_0 at time 0, the earnings next period are: $a_0(1 + CT^{\gamma-1})\exp(e - \overline{e} + \sigma\varepsilon - \sigma^2/2)$. The effects of talent, effort, and noise enter all multiplicatively. As in the Online Appendix to Gabaix and Landier (2008), the net present value of the CEO's action is proportional to the firm's market capitalization, to a very close first-order approximation.

[16] Cross-sectional variation in $\overline{e}_n$ 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).

where

$$\chi_n = \Lambda_n \overline{e}_n + \frac{\Gamma \Lambda_n^2 \sigma_n^2}{2} \tag{26}$$

denotes the "equivalent variation" associated with firm n, i.e. the utility loss suffered by the manager by exerting effort (the $\Lambda_n \overline{e}_n$ term) and bearing risk (the $\Gamma \Lambda_n^2 \sigma_n^2 / 2$ term). The latter arises because a fraction Λ_n is invested in the firm, which has volatility σ_n. After adjusting for the cost of effort and risk aversion, CEO n's "effective" wage is $\nu_n = w_n e^{-\chi_n}$.

As in Section 1.2, we derive the market equilibrium with a continuum of CEOs and a continuum of firms. To simplify the analysis, we assume that the firms' χ_n's are drawn independently of firm size.

Theorem 2 *(Pay and optimal incentive contract in market equilibrium). Let n_* denote the index of a reference firm. In equilibrium, the manager of rank i runs a firm whose "effective size" $e^{-\chi_n/\gamma} S$ is ranked i, and receives an expected pay:*

$$w_n = D(n_*)\, C S(n_*)^{\beta/\alpha} S_n^{\gamma-\beta/\alpha} \exp\left(\frac{\beta}{\alpha\gamma}(\chi_n - \overline{\chi})\right), \tag{27}$$

where $D(n_*) = -n_* T'(n_*) / (\alpha\gamma - \beta)$, χ_n is as defined in (26) and $\overline{\chi}$ is defined by $e^{-\overline{\chi}} = E\left[e^{-\widetilde{\chi}/(\alpha\gamma)}\right]^{\alpha\gamma}$ where $\widetilde{\chi}$ is the average of the firms' equivalent variations. The optimal contract is as given by Theorem 1, so that the final payoff W_n depends on the gross firm return R according to:

$$W_n = w_n \frac{R_n^{\Lambda}}{E\left[R_n^{\Lambda}\right]}, \tag{28}$$

where $E\left[R_n^{\Lambda}\right] = \exp\left(\sigma_n^2\left(\Lambda_n^2 - \Lambda_n\right)/2\right)$. As before, the wealth-performance sensitivity of CEO i is $\partial \ln W / \partial \ln R = \Lambda$, and the scaling with size are as in the basic model of Section 1.

To interpret Theorem 2, first note that the equivalent variation (26) χ_n increases in the cost of effort required by the firm ($\Lambda_n \overline{e}_n$) and firm risk (σ_n). A firm with higher equivalent variation χ_n will, ceteris paribus, choose a lower quality manager (since its effective size is $S_n e^{-\chi_n/\gamma}$), but with a higher pay. This is because the effective size $S_n e^{-\chi_n/\gamma}$ leads to a net wage $v_n \propto \left(S_n e^{-\chi_n/\gamma}\right)^{\gamma-\beta/\alpha}$, and a full wage $w_n = v_n e^{\chi_n} \propto S_n^{\gamma-\beta/\alpha} e^{\chi_n \beta/\alpha\gamma}$, which is increasing in χ_n. Therefore, 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, wages do not change. In equation (27), both χ_n and $\overline{\chi}$ increase by δ, so the wage is unchanged: even though working for his present firm becomes less attractive, the outside options also become less attractive.[17] To understand the result, for clarity we consider the case where all equivalent variations χ_n are the same, $\chi_n = \overline{\chi} = \chi$.

[17] This assumes that a CEO's only outside option is to become a CEO of another firm. If CEOs can find a job outside of the CEO market, the more general prediction is that the cross-sectional elasticity of wage to effort is higher than the market-wide elasticity.

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 (39):

$$w(n) = -\int_n^N CS(s)^{\gamma} T'(s)\, ds + \underline{w}_N e^{\chi n} \tag{29}$$

and scales according to

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

Changes in χ have negligible effect on the the pay of top CEOs, and zero effect in the limit as $n/N \to 0$. Equation (29) 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, and not affected at all in the asymptotic limit of top CEOs.

The main theoretical results of this paper – the determinants of incentives and total pay in market equilibrium – can be summarized in just three simple closed-form equations, (26)-(28).The wage depends on own firm size S_n, aggregate firm size $S(n_*)$, the supply of talent $D(n_*)$, the cost χ_n of effort and risk aversion that the firm imposes on the CEO, the market average of this cost, $\overline{\chi}$. Its incentive component is given by equation (28), an optimal unrestricted contract with a natural economic interpretation.

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 explained by optimal contracting.

3.1 Determinants of CEO Incentives

We start by examining the model's predictions for the cross-sectional scaling of incentive pay with firm size. These are summarized in Proposition 5 for the basic model, and are unchanged in the extended model. Our model predicts that the dollar-dollar wealth-performance sensitivity, B^{II}, should optimally decline with 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 $\xi \simeq 0.68$.[18] 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.[19] We calculate the wealth-performance sensitivities as follows:

$$B^{I} = \frac{1}{w_t}\left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \quad (30)$$

$$B^{II} = \frac{1}{S_t}\left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \quad (31)$$

$$B^{III} = \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right] \quad (32)$$

We use the Core and Guay (2002) methodology to estimate the option deltas. (Appendix C 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:[20]

$$\begin{aligned} \ln(B^{I}_{i,t}) &= \alpha + \beta \times \ln(S_{i,t}) \\ \ln(B^{II}_{i,t}) &= \alpha + \beta \times \ln(S_{i,t}) \\ \ln(B^{III}_{i,t}) &= \alpha + \beta \times \ln(S_{i,t}). \end{aligned}$$

Table 2 illustrates the results, which are consistent with the predictions of equation (15).

[18] 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.

[19] 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.

[20] We use the standard panel-data method which assumes the coefficients β are constant across firms. An alternative approach would be to allow β to vary between firms according to observed characteristics, as in Hermalin and Wallace (2001). They estimate the pay-performance relationship and that inter-firm differences will lead to this sensitivity differing between firms. Our focus here is instead the WPS-size relationship, and it is not clear that this will vary between firms. We therefore use the standard approach.

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%.

Insert Table 2 about here

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. Size independence permits meaningful comparisons of the strength of incentives across firms or over time. In regressions, it ensures that the explanatory power of the incentives measure does not simply arise because it proxies for size. If size is included separately as a regressor, it ensures that the coefficient on size is not distorted by the inclusion of another size proxy (i.e. incentives) on the right-hand side.

The extended model in Section 2 shows that the size scalings are unchanged when introducing risk aversion and general contracts. It also derives further predictions for the effect of cost of effort and risk on compensation. As predicted by equation (27), we find that the wage is significantly increasing in firm risk along the cross-section, with a coefficient of 0.61 on log volatility and a standard error of 0.09. This result is not reported in a table for brevity. Unfortunately, we cannot test the related predictions for the cost of effort as this measure cannot be quantified.

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 $B^I = \Lambda \frac{W}{w}$ (equation (19)). We present figures for 2001, the median year in our sample by level of incentives. The median B^I in 2001 is 9.[21]

We therefore calibrate $\Lambda = B^I w/W = 9w/W$. Shirking increases the CEO's utility by a fraction $\Lambda |\underline{e}| = 9\frac{w}{W} |\underline{e}|$of his wealth, i.e. $\$9w |\underline{e}|$ in dollar terms. $|\underline{e}|$ is the percentage amount by which CEO can reduce firm value by shirking or empire-building (through organic expansion or an acquisition). A natural starting point is the average takeover premium of 30%.[22] However, the takeover premium can be motivated by factors other than managerial misbehavior, such as synergies or undervaluation. Since a high input for $|\underline{e}|$ would make it easier to match the B^I found in the data, we conservatively set $\underline{e} \simeq 10\%$ which yields $\Lambda |\underline{e}| = \$0.9w$. The current level of incentive pay is able to deter actions for which the "private benefits of shirking" increase the CEO's utility by an amount no greater than 0.9 times his annual salary.

This appears a high upper bound which incorporates the majority of potential value-destructive actions, and so it may seem that observed incentives are able to address a number of agency issues. However, incentives are not effective in two cases: if the utility from shirking is very high, or the effect on the stock price is low. For certain actions, the private benefits from suboptimal behavior may exceed the upper bound. One example may be managerial entrenchment, as by failing to (optimally) resign, the manager retains his salary (plus private benefits of control) in many future years, the present value of which may plausibly exceed his annual pay.[23] Moreover, our estimate of $\$0.9w$ hinges upon our chosen input for $\underline{e}$ (it does not require an estimation of W/w, since it cancels out). For actions with smaller negative effects on the stock price, observed incentives will be too low to deter misbehavior even if it leads to modest private benefits. For example, if a managerialist acquisition or pet project only reduces stock returns by 1%, the manager will undertake it if the private benefits are greater than $\$0.09w$. We consider these cases in more detail in Section 4.1.

To calibrate Λ as a percentage of wealth, we would need to estimate W/w. Unfortunately, there is no data available on the wealth W of U.S. CEOs.[24] However, ExecuComp provides data on a CEO's financial wealth in his own firm. Across our sample, we estimate a median value of (Financial wealth in the firm) / (Pay) equal to 9.6. We assume that the CEO's wealth

[21] 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.

[22] Bennedsen, Perez-Gonzalez and Wolfenzon (2007) quantify the value lost from CEO distraction resulting from family deaths. Since distraction is not an example of wilful misbehavior, we base our calibrations on the takeover premium.

[23] Another example may be acquisitions that substantially boost firm size. Bebchuk and Grinstein (2007) find that increases in size lead to higher CEO pay in many future periods.

[24] We thank David Yermack for discussions on this point. See Becker (2006) for a study with Swedish CEOs.

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 38.4. We therefore have $\Lambda \left|\underline{e}\right| = 9\frac{w}{W}0.1 \simeq \frac{0.9}{38.4} = 0.23$.This means that, if the CEO shirks, his utility increases by an amount equivalent to 2.3% of his wealth.

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 by empiricists. For example, $B^{II} = B^{I}\frac{w}{S}$. The median size of the top 500 firms in 2001 is \$10 billion, with median pay of \$4.7 million. $B^{I} = 9$ is therefore consistent with a Jensen-Murphy semi-elasticity of $B^{II} = 9 \times$ (\$4.7 million) / (\$10 billion). This represents a wealth rise of \$4.23 for a \$1,000 increase in firm value, close to our directly measured figure of \$3.68.[25]

4 Extensions

This section considers extensions and other specifications of the model. Section 4.1 shows that actions that are additive in firm value, such as perk consumption, either cannot or should not be deterred by incentive compensation. In Section 4.2 we show that our model predicts a positive relationship between firm volatility and wealth volatility, which we support empirically. By contrast, traditional models that feature an unbounded effort domain have the opposite prediction. Section 4.3 shows that our multiplicative functional forms 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 4.4 reconciles our results with the empirical results of Baker and Hall (2004).

4.1 Perks

4.1.1 Perks in the Risk-Neutral Model

In the basic model, where the contract consists of cash and shares, the analysis assumed that $\Lambda < 1$, 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 $\Lambda > 1$ might apply is CEO entrenchment, since resigning adversely impacts the CEO's utility in many future periods. 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)).

[25] 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 median is \$8.37.

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. It also may hold for managerial rent extraction (e.g. stealing corporate resources).

Proposition 7 *(Impossibility of deterring perk consumption through incentive pay). Assume* $e = \underline{e}$ *reduces firm value by* $\$\overline{L}$*. 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 with simply cash and shares. Since the perk is fixed in absolute terms, the stock price of a large firm is relatively insensitive to perk consumption: stock returns only fall by $\overline{L}/S$. (The same holds for multiplicative actions for which $\underline{e}$ is small in magnitude, as considered earlier). 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 $w\Lambda\,|\underline{e}|$) 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 $f \geq 0$.

Although 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.[26] However, perks are typically independent of firm value, and thus cannot be addressed by incentives. As with the entrenchment issue, perks should instead be controlled by active corporate governance. For example, the board could intensely scrutinize the purchase of a corporate jet or a large investment project. Empirical evidence linking governance to shareholder returns (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 an important incremental role for governance, particularly in large firms.

Overall, these results show that incentives are effective in solving large agency problems, which have a significant effect on the stock price, but not smaller issues as these do not affect stock returns and thus the CEO's portfolio. However, these smaller issues are less important for

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

overall firm value. Any agency problem that would have a substantial effect on firm value also would have a substantial effect on stock returns, and so incentives are effective. Any agency problem that cannot be prevented by incentive compensation, because it has too small an effect on stock returns, is also less value-destructive if unchecked. Therefore, a greater problem may be an overconfident CEO. His actions may have significant negative effects on the stock price, yet incentives may be ineffective at deterring them as he genuinely believes that they are maximizing shareholder value.

4.1.2 Perks With Risk Aversion and Unrestricted Contracting

We now extend the above result to general incentive contracts. We will see that, although perk consumption can be deterred through the use of highly sensitive instruments, the required contract would impose substantial risk on the CEO that vastly outweighs the gains from deterring the perk. Direct control therefore remains the optimal method of perk prevention.

We use the optimal incentive scheme of Theorem 1, which states the optimal contact is to make the CEO invest a fraction $\theta = \Lambda$ of his wealth in the portfolio. 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 $(1-\lambda)\overline{L}$. The inefficiency of perk consumption is thus decreasing in λ, where $0 < \lambda < 1$. Perk consumption reduces stock returns by $\overline{L}/S$, so, if the CEO has a fraction θ of his wealth invested in the firm, the perk consumption reduces the value of his portfolio falls by $W\theta\overline{L}/S$. The CEO therefore avoids the perk if and only if $\lambda\overline{L} - W\theta\overline{L}/S \leq 0$, i.e. $\theta \geq \lambda S/W$. Therefore the optimal contract entails $\theta = \lambda S/W$.[27]

To illustrate the extreme sensitivity of incentives required, we consider a simple numerical example. If $\lambda = 1/2$, $S =$ \$10 billion and $W =$ \$100 million, perk prevention requires $\theta = 50$. This implies that the CEO must invest 5,000% of his wealth in the firm, borrowing to reach that amount (and continuously rebalancing, to maintain this exposure to firm return, and avoid personal bankruptcy). This is clearly extreme, and very costly for any non-trivial level of risk aversion. The reason for this high sensitivity is that consuming (say) \$10 million of perks for one year reduces the market capitalization by only 0.1%, an amount very hard to detect.

We now quantify the cost of inefficient risk-sharing and compare it to the benefits of perk prevention in a total surplus analysis. Since the CEO's utility is:

$$U = u\left(We^{-\Lambda\overline{e}-\Gamma\theta^2\sigma^2/2}\right),$$

the loss to total surplus 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

[27] This condition can also be derived using the framework of Section 2.1. Consuming perks $-e \in \left[0, \overline{L}\right]$ for a time Δt impacts CEO utility by $g(e) = \exp\left(-\frac{\lambda e}{W}\Delta t\right)$, and multiplies firm value by $L(e) = \exp\left(\frac{e}{S}\Delta t\right)$. Hence, the marginal cost of effort (24) is $\Lambda = \lambda S/W$, so the optimal portfolio share in Theorem 1 is $\Lambda = \lambda S/W$.

perk prevention if and only if

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

With $\Gamma = 1$ and $\sigma^2 = 0.04$ (an annual volatility of 20%), the right-hand side is equal to 1%. If the perk reduces firm value by \$10 million, 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. (The exception is for very small firms, where W is of a similar magnitude to S).

If the relative risk aversion of the CEO, Γ, is greater than 0.1, any perk less than \$1 billion ($\frac{\lambda^2}{1-\lambda}\frac{\Gamma\sigma^2}{2}\frac{S^2}{W}$, should not be deterred via incentives and can only be prevented through monitoring. If monitoring is not possible, then it is simply more cost-effective to let the CEO consume the perk, rather than try to deter it by incentives.

We summarize this result in the next Proposition:

Proposition 8 *(Perk prevention with general incentive contracts). Perks can be deterred with general incentive contracts if the CEO receives a share:*

$$\theta \geq \lambda\frac{S}{W}. \tag{33}$$

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

$$\overline{L} \geq \frac{\lambda^2}{1-\lambda}\frac{\Gamma\sigma^2}{2}\frac{S^2}{W}. \tag{34}$$

where S is firm size, W CEO wealth, Γ CEO relative risk aversion, σ the firm volatility, and λ the efficiency of perks.

While Proposition 7 could be achieved with simple partial equilibrium models, a calibratable framework is necessary to extend the result to general contracts (Proposition 8). It quantifies the cost of risk-bearing imposed by highly sensitive instruments, and shows that this exceeds the cost of the perk.

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

This section shows that the unbounded effort, featured by traditional models, leads to a predicted negative association between pay volatility and firm volatility. By contrast, our model has the opposite prediction, which we show to be empirically supported.

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

The solution is standard.[28] The CEO's dollar-dollar-pay-performance sensitivity is $b^{II} = \partial c/\partial S_1 = L/(L^2+a\sigma_r^2)$, 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 leads to optimal incentives being lower. By contrast, our model predicts that pay-performance sensitivity is independent of risk (see Section 1). There is a corner solution and no trade-off: since the firm (and thus the benefits of effort) is much larger than the manager (and thus the cost of effort in terms of risk-bearing), it is always efficient to implement the maximum level of effort. Indeed, Prendergast's (2002) survey of the evidence finds no systematic negative relationship between incentives and firm risk. He offers an explanation based on the allocation of responsibility to employees; ours is a complementary hypothesis.

In addition, models with bounded effort predict a negative relationship between pay volatility and firm volatility. Since pay volatility is $stdev(c) = \nu\sigma_r = \sigma_r SL/(L^2+a\sigma_r^2)$, its sensitivity to firm volatility is given by $\partial stdev(c)/\partial\sigma_r = -S\left(1-2b^{II}\right)b^{II}$. Since empirical studies find that b^{II} is substantially less than 1/2, these models predict $\partial stdev(c)/\partial\sigma_r < 0.$, i.e. that the CEO wealth volatility is smaller in very volatile firms.

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

$$stdev(W_{t+1}-W_t) = B^{III}\sigma_r \propto S^{\rho}\sigma_r, \tag{35}$$

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.[29] As discussed more fully in Appendix C, 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. $stdev(W_{t+1}-W_t) = B_t^{III}\sigma_r$.[30] The second uses ex post realized volatility, i.e. $stdev(W_{t+1}-W_t) = \ln|W_{t+1}-W_t|$.

[28] Normalizing the initial share price to $P=1$, the CEO's realized pay is $c = f+\nu(1+Le+\eta)$. The CEO chooses e to maximize his utility, $U = f+\nu(1+Le)-\frac{a}{2}\sigma_r^2\nu^2-\frac{1}{2}e^2$, and selects $e=\nu L$. The firm chooses ν to maximize its net value, $S\left(1+\nu L^2\right)-\frac{a}{2}\sigma_r^2\nu^2-\frac{\nu^2L^2}{2}$, and selects $\nu = SL^2/\left(L^2+a\sigma_r^2\right)$. The CEO's total pay is therefore $c = f+S_1L/\left(L^2+a\sigma_r^2\right)$.

[29] 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.

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

(We calculate wealth by starting with the CEO's initial holdings of stock and options and, each year, adding the appreciation in value of this portfolio plus any new flow compensation. We do not have data on the CEO's wealth outside of his firm.) 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 $\ln(|W_{t+1} - W_t| / w_t)$ and the model predicts $\beta_S = 0$ and $\beta_\sigma = 1$. Scaling by size yields $B_t^{II} \sigma_r$ or $\ln(|W_{t+1} - W_t| / S_t)$, with a prediction of $\beta_S = -2/3$ and $\beta_\sigma = 1$.

Insert Table 3 about here

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 $\ln(|W_{t+1} - W_t| / w_t)$ 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.

4.3 The Requirement for Multiplicative Preferences

Our choice of the multiplicative specification (1) was motivated by its intuitive plausibility, in particular that the benefits from shirking are increasing in the wage. Such a functional form generated the prediction that b^I is independent of w, which we have validated empirically. We now demonstrate that multiplicative preferences are necessary (as well as merely sufficient) to yield this implication.

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), which uses the form $E[c] - g(e)$. 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^I the fraction of w invested in stock, so that $c = w\left(1 + b^I r\right)$. As before, $b^I = E[\partial c/\partial r] / E[c]$. With the utility function $E[c] - g(e)$, the optimal b^I is given by $b^I = \frac{g(\underline{e}) - g(0)}{w}$, which implies:[31]

$$b^I \propto w^{-1} \tag{36}$$

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.

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

Another popular utility function is $E\left[c^{\alpha}/\alpha\right]-g\left(e\right)$, with $\alpha\in(0,1]$. 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 $e\in\{\underline{e},0\}$. Assume the firm's return is $r=e$ and that incentive compensation is implemented with shares, so the firm selects expected pay $\overline{c}$ and slope b^{I} so that: $c=\overline{c}\left(1+b^{I}r\right)$. The optimal contract minimizes $\overline{c}$ and b^{I} while granting the CEO his reservation utility of $\underline{u}$and eliciting $e=0$.[32] The next Proposition states that multiplicative preferences are required for the optimal $b^{I}=E\left[\partial c/\partial r\right]/E\left[c\right]$ to be independent of $\underline{v}$ (and thus $E[c]$).

Proposition 9 *(Necessity and sufficiency of multiplicative preferences to generate a size-independent b^{I}). Assume the CEO's utility function is $u\left(c,e\right)$, and the firm's return is $r=e$. Suppose the optimal affine contract involves a wage-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:*

$$u\left(c,e\right)=\phi\left(c\cdot g\left(e\right)\right) \tag{37}$$

for some functions ϕ and g.

Conversely, if preferences are of the type (37), then the optimal contract has a slope b^{I} that is independent of $E\left[c\right]$.

This result may be relevant for future calibratable models of corporate finance. 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.

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.[33]

4.4 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

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

[33] 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 $Ac^{1-\Gamma}/\left(1-\Gamma\right)+B$.

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.[34] 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),[35] 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 $\rho = 1/3$, 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.

5 Conclusion

This paper has presented a calibratable model of the competitive determination of the CEO compensation contract. There are two main theoretical contributions. First, it is a market equilibrium model endogenizing the level of total pay as well as its incentive component. As such, it is a unified framework for understanding the effect of many factors on the two main components of executive compensation. Second, it is particularly tractable and yields closed-form solutions, even when the model is extended to incorporate general incentive contracts. These features lead to clear empirical predictions and readily lend the model to empirical analysis. The main implications are as follows:

(i) Dollar-dollar incentives (such as those calculated by Jensen and Murphy (1990)) optimally decline with firm size, with an elasticity of -2/3. Therefore, the negative scaling observed empirically is fully consistent with optimal contracting and need not reflect inefficiency. Relatedly, dollar-log incentives should have a size elasticity of 1/3.

[34] 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.

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

(ii) 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.

(iii) Increased firm volatility is associated with increased wealth volatility, but does not affect the incentive component of total pay.

(iv) Higher firm risk and cost of effort lead to greater total pay in the cross-section, particularly for the least-talented CEOs. However, aggregate-level changes in these variables have no effect.

(v) Incentive compensation is typically effective at deterring value-destructive actions that have a large multiplicative effect on firm value. They are ineffective at preventing actions with a fixed dollar effect on firm value, particularly in large companies.

(vi) Observed levels of wealth-performance sensitivity are sufficient to deter value-destructive actions that yield private benefits no greater than 0.9 times the annual wage.

There are a large number of other potential determinants of compensation upon which the model is silent. Owing to its tractability and empirical consistency, our model may provide a useful benchmark on which future models can be built to explore 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), and renegotiation. In addition, there are a number of implications of the current model which we have not yet tested. 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?[36] 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?

One important caveat is that 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 as a partial justification of our assumptions, and in turn to support 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.

[36] 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).

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 when interpreting various features of the data.

A Detailed Proofs

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

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

The manager chooses $e = 0$ if:

$$E[cg(0) \mid e=0] \geq E[cg(\underline{e}) \mid e=\underline{e}].$$

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

$$w \geq \frac{w + \nu P\underline{e}}{1 - \Lambda\underline{e}} \Leftrightarrow \nu P \geq \nu^* P = w\Lambda.$$

f^* is chosen to ensure that expected pay is w: $f^* = w - \nu^* P = w(1-\Lambda)$.

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 $S(n)$ and manager $m \in [0, N]$ has talent $T(m)$. Low n denotes a larger firm and low m a more talented manager: $S'(n) < 0$, $T'(m) < 0$. n (m) 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

$$S' = S + CTS^\gamma, \tag{38}$$

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 $w(m)$ 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:

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

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

$$w(n) = -\int_n^N CS(u)^\gamma T'(u)\,du + \underline{w}_N. \tag{39}$$

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

$$w(n) = \int_n^N A^\gamma BCu^{-\alpha\gamma+\beta-1}du + \underline{w} = \frac{A^\gamma BC}{\alpha\gamma-\beta}\left[n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}\right] + \underline{w}_N \sim \frac{A^\gamma BC}{\alpha\gamma-\beta}n^{-(\alpha\gamma-\beta)}. \tag{40}$$

To interpret equation (40), 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 $w(n) = D(n_*)\,S(n_*)^{\beta/\alpha}S(n)^{\gamma-\beta/\alpha}$, where $S(n_*)$ is the size of the reference firm and $D(n_*) = -Cn_*T'(n_*)/(\alpha\gamma-\beta)$ is a constant independent of firm size.[37]

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

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

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

[37] The derivation is as follows. Since $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_*T'(n_*) = -Bn_*^\beta$, we can rewrite equation (40) as follows:

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

Proof of Theorem 1 For transparency, we first present a heuristic derivation, to demonstrate the essence behind equation (25). We then present the rigorous proof.

A heuristic proof. The IC condition is:

$$\overline{e} \in \arg\max_{e} c\left(\left(1+\eta\right)L\left(e\right)\right)g\left(e\right) \tag{41}$$

The derivative of the right-hand side of (41) at $e = \overline{e}$ should be non-negative i.e., using $R = (1+\eta) L(e)$:

$$c'(R)(1+\eta)L'(\overline{e})g(\overline{e}) + c(R)g'(\overline{e}) \geq 0,$$

i.e., using (24), $\Lambda = -\left(g'(\overline{e})/g(\overline{e})\right)/\left(L'(\overline{e})/L(\overline{e})\right)$:

$$c'(R) - \Lambda\frac{c(R)}{R} \geq 0. \tag{42}$$

Suppose that IC constraint (42) binds, which we will prove shortly. Then, $c'(R)R - \Lambda c(R) = 0$, which integrates to $c(R) = c_0 R^{\Lambda}$. C_0 is chosen to ensure that the participation constraint binds.

The above gives the "spirit" of why the Theorem holds. We now turn to a full proof.

The full proof. The problem is a hidden information problem. Using the revelation principle, after he learns η, the agent (the CEO) sends the principal (the firm) a message $\widehat{\eta}$ about the value of η. He then exerts effort e, and the return $R = (1+\eta)L(e)$ is realized. The firm optimizes over the optimal compensation contract, which is a function $c(\widehat{\eta}, R)$. The IC constraint is that the agent should report truthfully to the principal, and exerts maximal effort:

$$\forall\eta, \forall\widehat{\eta}, \forall e, u\left(c\left(\widehat{\eta}, (1+\eta)L(e)\right)g(e)\right) \leq u\left(c\left(\eta, (1+\eta)L(\overline{e})\right)g(\overline{e})\right)$$

i.e., as u is increasing,

$$\text{(IC) } \forall\eta, \forall\widehat{\eta}, \forall e, c\left(\widehat{\eta}, (1+\eta)L(e)\right)g(e) \leq c\left(\eta, (1+\eta)L(\overline{e})\right)g(\overline{e}). \tag{43}$$

u drops out, which is made possibly by the functional form (23).

The firm's problem is to minimize the expected cost $E\left[c\left(\eta, (1+\eta)L(\overline{e})\right)\right]$, subject to the IC constraint (43), and the reservation constraint $E\left[u\left(c\left(\eta, (1+\eta)L(\overline{e})\right)\right)\right] = u(\underline{w})$.

We observe that for any optimal contract, we can create a new optimal contract, replacing $c(\widehat{\eta}, R)$ by 0 if $R \neq (1+\widehat{\eta})L(\overline{e})$. The new contract still satisfies (43), and has the same cost. Hence, we restrict ourselves to contracts such that

$$c(\widehat{\eta}, R) = 0 \text{ if } R \neq (1+\widehat{\eta})L(\overline{e}). \tag{44}$$

Economically, this means that the principal pays 0 to the agent if he can infers that, from the agent's truthfully reported $\widehat{\eta}$ and the realized return, that the agent has not exerted maximum effort.

Now define $c(R) = c(R/L(\overline{e}) - 1, R)$. This is the consumption of an agent that exerts maximum effort and reports the true value of the noise η, yielding a full return $R = (1+\eta) L(\overline{e})$. Let us rewrite the IC condition (43) in terms of c. Owing to (44), given η, the relevant deviations in $(\widehat{\eta}, e)$ are only those such that $(1+\eta) L(e) = (1+\widehat{\eta}) L(\overline{e})$ So (43) can be rewritten as:

$$\text{(ICa): } \forall \eta, \forall e, c((1+\eta) L(e)) g(e) \leq c((1+\eta) L(\overline{e})) g(\overline{e}) \tag{45}$$

This is exactly (41) above. Hence, by the reasoning of the "heuristic proof", we have the necessary condition that $c'(R) - \Lambda \frac{c(R)}{R} \geq 0$.

We next prove that the IC constraint (42) binds. Let $p(1+\eta)$ denote the density function associated with $1+\eta$, and form the Hamiltonian H associated with $\min \int c(R) p(R) dR$ subject to $\int u(c(R)) p(R) dR \geq u(w)$ and $c'(R) - \Lambda c(R)/R \geq 0$:

$$H(R) = (c(R) + \mu \cdot u(c(R))) p(R) + h(R)(c'(R) - \Lambda c(R)/R).$$

We note that the problem is well defined, as it maximizes a linear function of $c(\cdot)$, $\int c(R) p(R) dR$, subject to $c(\cdot)$ belonging to a convex set, the set of functions c such that $\int u(c(R)) p(R) dR \geq u(w)$ and $c'(R) - \Lambda c(R)/R \geq 0$.

The first order condition for the optimality of the contract is the Euler-Lagrange equation: $\frac{\partial H}{\partial c(R)} - \frac{d}{dR} \frac{\partial H}{\partial c'(R)} = 0$, i.e.

$$(1 + \mu \cdot u'(c(R))) p(R) - h(R) \Lambda/R - h'(R) = 0$$

When $h(R) \neq 0$ over an interval, then (42) binds, and $c(R) = c_0 R^{\Lambda}$. If over an interval, $h(R) = 0$, then $1 + \mu \cdot u'(c(R)) = 0$, and $c(R)$ is constant. But then, over that interval, $c'(R) = 0$, and (42) implies $c(R) = 0$. This cannot be reconciled with $c(R) = c_0 R^{\Lambda}$ over the interval where $h(R) \neq 0$. So, $h(R)$ is never 0 over an interval, and so the contract is $c(R) = c_0 R^{\Lambda}$ for some $c_0 > 0$.

We finally prove that $c(R) = c_0 R^{\Lambda}$ implies the global IC constraint (45).

For a concave function F, we have $F(e) \leq F(\overline{e}) + F'(\overline{e})(e - \overline{e})$. Because $\ln L$ and $\ln g$ are concave, $\ln L + \Lambda \ln g$ is concave, and,

$$\begin{aligned} \Lambda \ln L(e) + g(e) &\leq \Lambda \ln L(\overline{e}) + \ln g(\overline{e}) + \left[\Lambda (\ln L(\overline{e}))' + (\ln g(\overline{e})')\right](e - \overline{e}) \\ &= \Lambda \ln L(\overline{e}) + \ln g(\overline{e}) \text{ by (24).} \end{aligned}$$

Hence, for any e, $L(e)^{\Lambda} g(e) \leq L(\overline{e})^{\Lambda} g(\overline{e})$, and

$$\forall \eta, \forall e, c_0 (1+\eta)^{\Lambda} L(e)^{\Lambda} g(e) \leq c_0 (1+\eta)^{\Lambda} L(\overline{e})^{\Lambda} g(\overline{e}),$$

i.e. *(45)*. The proof is now complete.

Lemma 1 *When the firm S is large enough, the CEO should optimally exert high effort.*

Proof: Call $K(e)$ the dollar cost to make the CEO exert an effort level e, subject to his participation constraint $E[u(c)] \geq u(\underline{w})$. Under quite general conditions, $K(e)$ is continuously differentiable in $[\underline{e}, \overline{e}]$.

For instance, with the contract as in Proposition 1, $c = c_0 R^{\Lambda(e)}$, where $\Lambda(e) = -(\ln g(e))' / (\ln L(e))'$. The participation constraint is satisfied with $E\left[u\left(c_0 R^{\Lambda(e)}\right)\right] = w$, which admits a solution as $\lim_{c\to+\infty} u(c) = +\infty$. Then, $K(e) = c_0 E\left[R^{\Lambda(e)}\right]$. By the implicit function theorem, $K(e)$ is continuously differentiable.

The firm's surplus, net of compensation, is $V(e) = SL(e) - K(e)$. The firm solves $\max_e V(e)$. Since $V'(e) = SL'(e) - K'(e)$, we have $V'(e) > 0$ for $e \in [\underline{e}, \overline{e}]$, if $S > S^* = \max_e K'(e) / L'(e)$. Hence, for $S > S^*$, the firm wishes to implement the maximum level of effort.

We finally have to check that, conditional on a realization of η, the firm does wish to implement maximal effort. This is the case if, for all realizations of η, $\overline{e}$ maximizes the net surplus: $S(1+\eta) L(e) - W_0 ((1+\eta) L(e))^{\Lambda} g(e)$. This is true if S is large enough, and $\ln(1+\eta)$ has bounded support.

Proof of Theorem 2 Assume that in market equilibrium, a CEO of talent $T(m)$ 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 $v(m) e^{\chi_n}$. So its program is: $\max_m CS(n)^{\gamma} T(m) - v(m) e^{\chi_n}$, 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" $(e^{-\chi_n})^{1/\gamma} S(n)$ and thus pays the associated wage $v = D(n_*) (e^{-\overline{\chi}} S(n_*))^{\beta/\alpha} \left(C_n e^{-\chi_n/\gamma} S\right)^{\gamma-\beta/\alpha}$. After the compensating differential, the dollar wage is: $w = v e^{\chi_n}$, hence (27).

Proof of Proposition 9 Define $\phi(c) = u(c, 0)$, $g(0) = 1$ and $g(\underline{e}) = 1/\left(1 + b^I \underline{e}\right)$. Call $b^I = E[\partial c / \partial r] / E[c]$ the slope. Since b^I offers the minimum slope, $E[v(c,e) \mid e = 0] = E[u(c,e) \mid e = \underline{e}]$, i.e.

$$u\left(\overline{c}\left(1 + b^I \underline{e}\right), \underline{e}\right) = u(\overline{c}, 0) = \phi(\overline{c})$$

and so $u(c, \underline{e}) = \phi\left(c / \left(1 + b^I \underline{e}\right)\right) = \phi(c g(\underline{e}))$. Therefore, $u(c, e) = \phi(c g(e))$ for all c and $e \in \{\underline{e}, 0\}$.

The converse of the proof is immediate, with $b^I = (1 - g(0) / g(\underline{e}))$.

B Multiperiod Model

This Appendix underpins Section 1.4, which extends the pay-performance sensitivity results of Sections 1.1-1.3 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.[38] Let the value function V_t denote the discounted utility of future consumption:

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

For instance, if consumption and effort are deterministic, $\ln V_t = \sum_{s=0}^{\infty} \delta^s \left((1-\delta)\ln c_{t+s} - \Lambda e_{t+s}\right)$.[39]

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 $r_{t+1} = r_f + e_t + \eta_{t+1}$, where r is the risk-free rate and $e_t \in \{\underline{e}, 0\}$. Wealth evolves according to:

$$W_{t+1} = W_t\left(1 + r_f + \theta_t e_t + \theta_t \eta_{t+1}\right) - c_{t+1}. \tag{46}$$

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: $\Delta W_t = -W_r(t)\Delta t$,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:

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

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

$$\frac{\partial W}{\partial r} \geq \Lambda W \tag{47}$$

[38] 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.

[39] 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 $\Lambda \Delta t$.

As in Section 1, we select the contract that minimizes the risk in the CEO's pay. It is given by

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

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

While equation (47) makes 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 $\frac{\partial W}{\partial r} \geq \Lambda W$ so that incentive compatibility is achieved. The CEO's new wealth is $W' = W + W^{\Delta}$. To maintain incentive compatibility, we require $\frac{\partial W^{\Delta}}{\partial r} \geq \Lambda W^{\Delta}$., and so $\frac{\partial W'}{\partial r} \geq \Lambda W'$. The least risky contract satisfying this condition is given by:

$$\frac{\partial W^{\Delta}}{\partial r} = \Lambda W^{\Delta}. \tag{48}$$

The one-period model of Section 1 predicted exactly (48). 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 1, in particular Propositions 3, 4, and 5.

C 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 a 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 (2002). 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 intrinsic value 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

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

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 are 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

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

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:

$$\frac{W_{t+1} - W_t}{w_t} = A + \widehat{B^I} \times r_{t+1} + C \times r_{M,t+1} + \text{Controls}, \tag{49}$$

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).[40] This compares with our chosen measure of:

$$B^{I,\text{ex ante}} = \frac{1}{w_t}\left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P\right], \tag{50}$$

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 (49) 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:

$$\ln w_{t+1} - \ln w_t = a + \widehat{b^I} \times r_{t+1} + \text{Controls}, \tag{51}$$

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 $2\left(w_{t+1} - w_t\right)/\left(w_{t+1} + w_t\right)$, which is more robust when w_t is close to 0.

[40] $r_{M,t+1}$ 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 (49) 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.

References

[1] Almazan, A. and J. Suarez, 2003, “Entrenchment and Severance Pay in Optimal Governance Structures.” *Journal of Finance* 58, 519-548

[2] Armstrong, C., D. F. Larcker and C. Su, 2007, “Stock Options and Chief Executive Officer Compensation.” Working Paper, Stanford University

[3] Baker, G., M. Jensen and K. J. Murphy, 1988, “Compensation and Incentives: Practice vs. Theory.” *Journal of Finance* 43, 593-616

[4] Baker, G. and B. Hall, 2004, “CEO Incentives and Firm Size.” *Journal of Labor Economics* 22, 767-98

[5] Baranchuk, N., G. MacDonald and J. Yang, 2007, “The Economics of Super Managers.” Working Paper, University of Texas at Dallas

[6] Bebchuk, L. and J. Fried, 2004, *Pay Without Performance: The Unfulfilled Promise of Executive Compensation.* Cambridge, MA: Harvard University Press

[7] Bebchuk, L., and Y. Grinstein, 2005, “The Growth of Executive Pay.” *Oxford Review of Economic Policy* 21, 283-303

[8] Bebchuk, L., and Y. Grinstein, 2007, “Firm Expansion and CEO Pay,” Working Paper, Cornell University

[9] Becker, B., 2006, “Wealth and Executive Compensation.” *Journal of Finance* 61, 379-397

[10] Bennedsen, M. F. Perez-Gonzalez, and D. Wolfenzon, 2007, “Do CEOs Matter?” Working Paper, Copenhagen Business School

[11] Bertrand, M. and S. Mullainathan, 2001, “Are CEOs Rewarded for Luck? The Ones Without Principals Are.” *Quarterly Journal of Economics* 116, 901–932

[12] Core, J. and W. Guay, 1999, “The Use of Equity Grants to Manage Optimal Equity Incentive Levels.” *Journal of Accounting and Economics* 28, 151-184

[13] Core, J. and W. Guay, 2002, “Estimating the Value of Employee Stock Option Portfolios and Their Sensitivities to Price and Volatility”. *Journal of Accounting Research* 40, 613-630

[14] Core, J., W. Guay and R. Verrecchia, 2003, “Price Versus Non-Price Performance Measures in Optimal CEO Compensation Contracts.” *The Accounting Review* 78, 957-981

[15] Demsetz, H. and K. Lehn, 1985, “The Structure of Corporate Ownership: Causes and Consequences.” *Journal of Political Economy* 93, 1155-1177

[16] Dittmann, I. and E. Maug, 2007, "Lower Salaries and No Options? On the Optimal Structure of Executive Pay, " *Journal of Finance*, 62, 303-343

[17] Edmans, A., 2007, "Inside Debt". Working Paper, University of Pennsylvania

[18] Epstein, L. and S. Zin, 1989, "Substitution, Risk Aversion and the Temporal Behavior of Consumption and Asset Returns: A Theoretical Framework." *Econometrica* 57, 937-968

[19] Falato, A., and D. Kadyrzhanova, 2007, "Optimal CEO Incentives and Industry Dynamics." Working Paper, Federal Reserve Board

[20] Frydman, C. and R. Saks, 2007, "Historical Trends in Executive Compensation, 1936-2003." Working Paper, Massachusetts Institute of Technology

[21] Gabaix, X. and A. Landier, 2008, "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics*, forthcoming.

[22] Garicano, L. and T. Hubbard, 2005, "Learning About the Nature of Production from Equilibrium Assignment Patterns." Working Paper, University of Chicago

[23] Gayle, G. and R. Miller, 2007, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" Working Paper, Carnegie-Mellon University

[24] Gibbons, R. and K. J. Murphy, 1992, "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100, 468-505

[25] Gompers, P., J. Ishii and A. Metrick, 2003, "Corporate Governance and Equity Prices." *Quarterly Journal of Economics* 118, 107-155

[26] Grossman, S. J. and O. D. Hart, 1983, "An Analysis of the Principal-Agent Problem," *Econometrica* 51, 7-45

[27] Hall, B. and J. Liebman, 1998, "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 113, 653-691

[28] Haubrich, J. G., 1994, "Risk Aversion, Performance Pay, and the Principal-Agent Problem." *Journal of Political Economy* 102, 258-276.

[29] He, Z., 2007, "Optimal Executive Compensation When Cash Flows Follow Geometric Brownian Motion," *Review of Financial Studies*, forthcoming

[30] Hermalin, B. and M. Weisbach, 1998, "Endogenously Chosen Boards of Directors and Their Monitoring of Management." *American Economic Review* 88, 96-118

[31] Hermalin, B. and N. Wallace, 2001, "Firm Performance and Executive Compensation in the Savings and Loan Industry," *Journal of Financial Economics* 61, 139–170

[32] Himmelberg, C. and G. Hubbard, 2000, "Incentive Pay and the Market for CEOs: An Analysis of Pay-for-Performance Sensitivity." Working Paper, Columbia University

[33] Holmstrom, B., 1979, "Moral Hazard and Observability." *Bell Journal of Economics* 10, 74-91

[34] Holmstrom, B., 1992, "Comments," in Lars Werin and Hans Wijkander, eds., *Contract Economics*. Blackwell, Cambridge MA

[35] Holmstrom, B., 1999, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies,* 66, 169–182.

[36] Holmstrom, B. and S. Kaplan, 2001, "Corporate Governance and Takeovers in the US: Making Sense of the '80s and '90s." *Journal of Economic Perspectives* 15, 121-144

[37] Holmstrom, B. and P. Milgrom, 1987, "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55, 308-328

[38] Inderst, R. and H. Mueller, 2007, "CEO Compensation and Private Information: An Optimal Contracting Perspective." Working Paper, New York University

[39] Jensen, M. and W. Meckling, 1976, "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure." *Journal of Financial Economics* 3, 305-360

[40] Jensen, M. and K. J. Murphy, 1990, "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98, 225-64

[41] Jensen, M. and K. 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

[42] Kreps, D. and E. Porteus, 1978, "Temporal Resolution of Uncertainty and Dynamic Choice Theory." *Econometrica* 46, 185-200.

[43] Kuhnen, C. and J. Zwiebel, 2007, "Executive Pay, Hidden Compensation and Managerial Entrenchment." Working Paper, Northwestern University

[44] Lazear, E.P., 1995, *Personnel Economics*, MIT Press.

[45] Manso, G., 2006, "Motivating Innovation." Working Paper, Massachusetts Institute of Technology

[46] Morck, R., A. Shleifer and R. Vishny, 1990, "Do Managerial Objectives Drive Bad Acquisitions?" *Rand Journal of Economics* 24, 357-379

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

[48] Murphy, K. 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

[49] Oyer, P., 2004, "Why Do Firms Use Incentives That Have No Incentive Effects?," *The Journal of Finance*, 59, 1619-1649.

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

[51] Rajan, R. and J. Wulf, 2006, "Are Perks Purely Managerial Excess?" *Journal of Financial Economics* 79, 1-33

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

[53] Sannikov, Y., 2006, "A Continuous-Time Version of the Principal-Agent Problem." Working Paper, New York University

[54] Sattinger, M., 1993, "Assignment Models of the Distribution of Earnings," *Journal of Economic Literature* 31, 831–880

[55] Schaefer, S., 1998, "The Dependence of Pay-Performance Sensitivity on the Size of the Firm." *Review of Economics and Statistics*, 80, 436-443

[56] Tervio, M., 2007, "The Difference that CEOs Make: An Assignment Model Approach." Working Paper, University of California, Berkeley

[57] Weil, P., 1990, "Nonexpected Utility in Macroeconomics." *Quarterly Journal of Economics* 105, 29-42

[58] Yermack, D., 1995, "Do Corporations Award CEO Stock Options Effectively?" *Journal of Financial Economics* 39, 237-269

[59] Yermack, D., 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	$\frac{\Delta \ln c}{\Delta \ln S}$	$\frac{\Delta c}{\Delta S}$	$\frac{\Delta c}{\Delta \ln S}$
Real variables	$\frac{\text{\$shares}}{\text{total pay}}$	% shares	\$ shares
WPS analog	$\frac{\Delta \$W}{\Delta \ln S}\frac{1}{w}$	$\frac{\Delta \$W}{\Delta \$S}$	$\frac{\Delta \$W}{\Delta \ln S}$
Used by	Murphy (1985)	Demsetz-Lehn (1985)	Holmstrom (1992)
	Rosen (1992)	Jensen-Murphy (1990)	Hall-Liebman (1998)
		Yermack (1995)	
		Schaefer (1998)	
This paper	Λ	$\Lambda \frac{w}{S}$	Λw
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 three 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, W is the wealth, Λ is the cost of effort. 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, and $b^I \propto S^0$ means that b^I is constant across firm sizes.

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 (2002) methodology to estimate the delta of the CEO's option holdings. B^I, B^{II} and B^{III} are estimated using equations (30)-(32). 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		
	$\ln(B^I\sigma_r)$	$\ln(B^{II}\sigma_r)$	$\ln(B^{III}\sigma_r)$	$\ln\left(\frac{\lvert W_{t+1}-W_t\rvert}{w_t}\right)$	$\ln\left(\frac{\lvert W_{t+1}-W_t\rvert}{S_t}\right)$	$\ln\lvert W_{t+1}-W_t\rvert$
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 (2002) methodology to estimate the delta of the CEO's option holdings. B^I, B^{II} and B^{III} are estimated using equations (30)-(32). 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.