



2009


A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium

Alex Edmans
University of Pennsylvania

Xavier Gabaix

Augustin Landier

Follow this and additional works at: https://repository.upenn.edu/fnce_papers

 Part of the [Economic Theory Commons](#), [Finance Commons](#), and the [Finance and Financial Management Commons](#)

Recommended Citation

Edmans, A., Gabaix, X., & Landier, A. (2009). A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium. *Review of Financial Studies*, 22 (12), 4881-4917. <http://dx.doi.org/10.1093/rfs/hhn117>

This paper is posted at ScholarlyCommons. https://repository.upenn.edu/fnce_papers/330
For more information, please contact repository@pobox.upenn.edu.

A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium

Abstract

This paper presents a unified theory of both the level and sensitivity of pay in competitive market equilibrium, by embedding a moral hazard problem into a talent assignment model. By considering multiplicative specifications for the CEO's utility and production functions, we generate a number of different results from traditional additive models. First, both the CEO's low fractional ownership (the Jensen–Murphy incentives measure) and its negative relationship with firm size can be quantitatively reconciled with optimal contracting, and thus need not reflect rent extraction. Second, the dollar change in wealth for a percentage change in firm value, divided by annual pay, is independent of firm size, and therefore a desirable empirical measure of incentives. Third, incentive pay is effective at solving agency problems with multiplicative impacts on firm value, such as strategy choice. However, additive issues such as perk consumption are best addressed through direct monitoring.

Disciplines

Economic Theory | Finance | Finance and Financial Management

A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium*

Alex Edmans
The Wharton School

Xavier Gabaix
NYU Stern and NBER

Augustin Landier
NYU Stern

April 12, 2008

Abstract

Existing compensation models typically assume that effort has additive effects on CEO utility. This paper considers multiplicative specifications for the principal-agent problem, and further embeds the problem into a talent assignment model. The result is a unified framework endogenizing both incentives and total pay levels in competitive market equilibrium. The predictions generated by multiplicative specifications match a number of stylized facts inconsistent with an additive model. First, the negative relationship between the CEO's effective equity stake and firm size can be quantitatively explained by an optimal contracting model and thus need not reflect rent extraction. Second, our multiplicative setting predicts that the dollar change in wealth for a percentage change in firm value, scaled by annual pay, is independent of firm size and thus a desirable empirical measure. This independence is confirmed in the data. Third, incentive compensation is effective at solving large agency problems, such as strategy choice, but smaller issues such as perk consumption are best addressed through direct monitoring.

KEYWORDS: Executive compensation, multiplicative preferences, pay-performance sensitivity, incentives, perks, optimal contracting, calibration

JEL CLASSIFICATION: D2, D3, G34, J3

*aedmans@wharton.upenn.edu, xgabaix@stern.nyu.edu, alandier@stern.nyu.edu. This paper was previously circulated under the title "A Calibratable Model of Optimal CEO Incentives in Market Equilibrium." For helpful comments, we thank two anonymous referees, the editor (Michael Weisbach), Franklin Allen, Yakov Amihud, Indraneel Chakraborty, John Core, Ingolf Dittmann, Carola Frydman, Bob Gibbons, Xavier Giroud, Itay Goldstein, Zhiguo He, Ben Hermalin, Dirk Jenter, Holger Mueller, Derek Neal, Luis Palacios, Michael Roberts, Yuliy Sannikov, Merih Sevilir, Andrei Shleifer, Jeremy Stein, Youngsuk Yook and seminar participants at the University of Delaware, MIT, NYU, the Econometric Society Winter Meetings, the NBER, and the Washington University Conference on Corporate Finance. XG thanks the NSF for financial support.

This paper presents a neoclassical model for CEO incentives and total pay, which yields an optimal contracting benchmark against which current practices can be evaluated. Our approach features two main departures from existing compensation models. First, motivated by first principles and intuitive plausibility, we introduce multiplicative preferences into the principal-agent problem. The resulting empirical predictions match a number of stylized facts that traditional additive models find difficult to reconcile with optimal contracting. Second, while many existing models are partial equilibrium, taking the level of pay as given and focusing on its optimal division into fixed and performance-sensitive components, we endogenize total pay in general equilibrium by embedding the principal-agent problem into a competitive assignment model of CEO talent. The result is a parsimonious, unified model of incentives and total pay, where both components of compensation are simultaneously and endogenously determined by the market for scarce talent and the nature of the agency conflict. The framework is tractable and yields closed-form solutions. These give rise to testable predictions, which we validate empirically.

The first departure is our multiplicative specification for the costs and benefits of effort, which contrasts with the linear functional forms commonly used. With multiplicative preferences, the utility gain from shirking and private benefits is proportional to the CEO's wage. The model thus treats private benefits as a normal good, consistent with the treatment of most goods and services in consumer theory. The share of total pay allocated to consumption and leisure is independent of total salary, and so labor supply does not diverge over time as wages change. This empirical consistency explains the common use of multiplicative preferences in calibrated macroeconomic models (see, e.g., Cooley and Prescott (1995)).

With a multiplicative production function, effort has a percentage effect on firm value and so the dollar benefits of working are higher for larger firms. This assumption is plausible for the majority of CEO actions which can be "rolled out" across the entire firm, and thus have a greater effect in a larger company. For example, if the CEO designs a new method to reduce production costs, this can be applied firmwide. Similarly, strategic choice or the launch of new projects also affect the whole firm.

If the production function is linear, effort has little effect in large firms. Thus it is optimal to implement an interior level of effort, to avoid exerting excessive costs on the manager (disutility from effort plus risk-bearing). In our multiplicative setting, under quite weak assumptions, it is always efficient to induce maximum effort. This occurs because the cost of incentives is a function of the CEO's wage, but the benefits of effort are a function of firm value, which is substantially greater. Because maximum effort is always optimal, the efficient contract takes a simple form. Since effort has a percentage effect on both firm value and utility, the percentage change in pay for a percentage firm return is the relevant incentive measure, and it must be sufficiently high to induce maximum effort. Translated into real variables, this measure equals

the proportion of total salary that is comprised of shares.¹ If the CEO’s salary doubles, the dollar benefits of shirking also double. His dollar equity stake must also double to maintain incentive compatibility. Thus, the fraction of pay that must be composed of equity should be constant across CEOs of different salaries.

By contrast, in an additive model, effort has a fixed dollar effect on firm value and managerial utility, and so the dollar change in pay for a dollar increase in firm value is the appropriate measure (see, e.g., Jensen and Meckling (1976)). “Dollar-dollar”, rather than “percent-percent” incentives, are relevant. In real variables, the former is the CEO’s percentage equity stake in the firm, and linear models predict that this should be constant across CEOs.

The above contract only gives equity compensation as a fraction of total pay. Our second modeling contribution is to embed this principal-agent problem into general equilibrium to endogenize total pay, allowing us to fully solve for the absolute level of incentives and generate empirical predictions. We use the competitive talent assignment model of Gabaix and Landier (2008), where the most skilled CEOs are matched with the largest firms and earn the highest salaries. Since total pay varies with firm size, our model generates predictions for the relationship between incentives and firm size under first-best contracting. Note that Gabaix and Landier do not consider agency problems and thus make no predictions for CEO incentives.

The relationship between incentives and size is important for at least two reasons. It has been widely documented that the CEO’s effective equity stake (“dollar-dollar” incentives) is significantly decreasing in firm size (Demsetz and Lehn (1985), Jensen and Murphy (1990), Gibbons and Murphy (1992), Schaefer (1998), Hall and Liebman (1998) and Baker and Hall (2004)). As stated above, linear models predict that dollar-dollar incentives should be constant across CEOs and thus independent of size. One interpretation of this inconsistency between optimal contracting theory and observed practice is that incentives are inefficiently low in large firms, perhaps because governance is particularly weak in such companies (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 incentive determination is broken, and must be fixed.

Our model has the opposite conclusion. With a multiplicative production function, the dollar increase in firm value from CEO effort is proportional to size, i.e. has a elasticity of 1 with firm size. With multiplicative preferences, the CEO’s dollar utility gain from shirking rises with the wage, but wages only have a 1/3 elasticity with size (see Gabaix and Landier (2008) for a survey of the empirical evidence). Therefore, dollar-dollar incentives should have a size elasticity of $1/3 - 1 = -2/3$, which is very close to our empirical estimate of -0.60 (with

¹The optimal contract can also be implemented with other equity-like instruments, such as options and bonuses. Our contract gives the optimal amount of “share-equivalents”, where other instruments are converted to shares according to their deltas. See Section C for an extension of the model to general incentive contracts and Section D for detail on the empirical conversion of options.

a standard error of 0.05).

The observed negative relationship is therefore quantitatively consistent with optimal contracting. Simply put, since effort has such a high dollar effect in large firms, the manager will work even if he has a relatively small equity stake. 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.

While our choice of a multiplicative functional form was motivated by its intuitive plausibility and use in macroeconomics, rather than the desire to match moments, we then show that it is not only sufficient to match the empirical scaling, but also necessary. This result has implications for future quantitative models of CEO compensation: the desire for empirical consistency limits the functional form that can be used.

Understanding the scaling of incentive measures with firm size is also important to evaluate the various metrics available to empiricists. Our multiplicative model advocates a new empirical measure of CEO incentives: it suggests that “percent-percent” incentives are independent of firm size, a fact confirmed by the data. Translated into real variables, and allowing for CEO incentives to stem from existing holdings of equity as well as new flows, this measure is the “scaled wealth-performance sensitivity”: the dollar change in wealth for a percentage change in firm value, divided by annual pay. By contrast, existing commonly used measures vary strongly with firm size. Size invariance is a desirable property for an empirical measure, as it leads to comparability across firms and over time.

A second empirical prediction is the level of incentives. 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. Hall and Liebman (1998) disagree, arguing that “dollar-dollar” incentives are not the relevant measure of incentive compatibility. Indeed, we find that observed incentives are sufficient to deter suboptimal actions (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 many value-destructive actions with multiplicative impacts fall below this upper bound, incentives are able to solve many multiplicative agency problems. Again, our specification is central for this result: the cost of effort is proportional to CEO wealth and its benefit is proportional to firm value. Since the latter is substantially greater, the dollar gains from effort are very high and so even a small equity stake (i.e. small dollar-dollar incentives) will induce maximum effort. Haubrich (1994) identifies the parameter values in the linear model that would be consistent with Jensen and Murphy’s statistic. He notes that the large number of free variables (including risk aversion) makes it relatively easy to match one moment. Our model, which lacks a risk aversion parameter, can explain both the level of incentives and their

scaling with firm size.

A third stylized fact is the positive correlation between firm volatility and wealth volatility. Traditional models predict a negative relationship: higher firm volatility increases the risk-bearing costs imposed on the manager by incentive compensation. The optimal level of incentives, and thus effort, is lower. In this paper, as noted previously, it is always optimal to implement the maximum effort level, regardless of costs to the manager and thus volatility. Optimal incentives are independent of volatility; since wealth volatility equals the product of incentives and firm risk, the model generates the positive relationship found in the data.

We extend the model by noting that the multiplicative production function does not apply to all CEO decisions. 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 an additive effect. Since such actions have a very small effect on the equity returns of a large company, we show that no amount of equity compensation can deter perks. While the seminal model of Jensen and Meckling (1976) implies that all agency issues can and should be solved by incentives, we show that equity can only address large agency problems with a multiplicative effect on firm value. Smaller, “additive” issues such as perk consumption should instead be addressed through direct monitoring, and thus have no explanatory power for incentive compensation.

This paper is closely related to a number of recent structural models and calibrations of the CEO incentive problem. Dittmann and Maug (2007), Dittmann, Maug and Spalt (2008) and Armstrong, Larcker and Su (2007) explore the optimal structure of compensation, in particular the mix of stock and options. Garicano and Hubbard (2007) 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. The linear model closest to explaining the observed scaling between incentives and firm size is Baker and Hall (2004). Theirs is an “inversion” model, which assumes observed incentives are efficient and “backs out” the production function that would be consistent. By contrast, our paper motivates specifications from first principles, and then compares the resulting predictions with the data to evaluate the efficiency of current practices. In addition, it considers the effect of preferences as well as production functions on incentives. Like Baker and Hall, Coles, Lemmon and Meschke (2007) also use incentives as an input. They estimate the productivity of both managerial effort and physical capital, and in turn use these parameters to generate the stylized U-shaped relationship between Tobin’s q and managerial ownership.

Our paper differs from the above theories owing to its contrasting objectives (principally, the level and scaling of incentives²) and its modeling approach (multiplicative specifications and a general equilibrium approach incorporating both pay and incentives.) Baranchuk, Macdonald

²Dicks (2007) predicts a negative relationship between incentive pay and firm size through a different channel: governance is stronger in large firms, reducing the need for monetary incentives. He (2007) also finds a negative relationship with geometric Brownian cash flows and CARA utility. In our paper, the CEO is risk neutral.

and Yang (2007) and Falato and Kadyrzhanova (2007) also present general equilibrium models of incentives, although without multiplicative functional forms and with different purposes. The former endogenizes firm size and focuses on the effect of product market conditions on CEO compensation. The latter analyzes the effect of industry competition and a firm’s competitive position on optimal contracts.

A separate literature to which this paper relates examines the optimality of CEO compensation practices. Our paper focuses on the level and scaling of incentives, but there are a large number of other stylized facts of the CEO labor market not considered by our model and which may indeed result from rent extraction (Bebchuk and Fried (2004)). Examples include the widespread use of “hidden” compensation, the lack of relative performance evaluation, the high pay of U.S. CEOs compared to the rest of the world, the widespread use of at-the-money options, and positive market reactions to deaths of a potentially optimally contracted CEO. Our model’s tractability and empirical consistency may render it a benchmark upon which future theories can build, to explore some of the issues on which the current paper is silent while continuing to match the level and scaling of incentives. In Edmans and Gabaix (2008) we extend the current framework to incorporate risk aversion and general contracting instruments under continuous time, showing that the key results remain robust and generating additional predictions. A number of other theories investigate whether or not the above features can be consistent with efficiency, using different frameworks as they do not simultaneously attempt to match empirical incentives. 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)), perks (Rajan and Wulf (2006)), and the lack of indexation (Oyer (2004)). Kuhnen and Zwiebel (2007) model hidden compensation as inefficient rent extraction and show that a suboptimal contracting model can explain many features of the data.

This paper is organized as follows. In Section 1 we present our general equilibrium model with multiplicative functional forms, and derive empirical implications. Section 2 shows that these predictions quantitatively match the data. Section 3 considers further implications of the model, in particular the ineffectiveness of incentives at deterring perks, and Section 4 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 which endogenizes total pay. This leads to empirical predictions for the scaling of pay-performance sensitivity with firm size, detailed 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. Section 1.5 proves that multiplicative preferences are

not only sufficient, but also necessary to generate our scaling predictions.

Since our objective is to provide testable predictions, we maximize tractability by building a deliberately parsimonious model where the manager is risk-neutral, the effort decision is binary and the contract is restricted to comprise cash and shares. We show that our results are robust to multiple effort levels and general incentive contracts in Section 3.2 and Appendix C respectively. Owing to risk neutrality, there is a continuum of incentive-compatible contracts; we select the one that minimizes the variable component of compensation as this would be strictly optimal under any level of risk aversion. In Edmans and Gabaix (2008), we show that the model’s key results hold under a general risk-averse utility function and in continuous time.

1.1 Incentive Pay in Partial Equilibrium

The CEO’s objective function is:

$$U = E [cg(e)], \tag{1}$$

where c is the CEO’s monetary compensation and $e \in \{\underline{e}, \bar{e}\}$ denotes CEO effort. We normalize $\underline{e} = 0 < \bar{e} < 1$, and $g(\bar{e}) = 1 < g(0) = 1/(1 - \Lambda\bar{e})$, where $0 < \Lambda\bar{e} < 1$.

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, $e = \bar{e}$ represents high effort and $e = \underline{e}$ is shirking. A second interpretation is the choice of an investment project, strategy or acquisition target, where $e = \bar{e}$ is the first best project and $e = \underline{e}$ yields the CEO private benefits, such as an empire-building expansion. We use the term “shirking” and “private benefits” interchangeably. Shirking increases the CEO’s utility by (approximately) a fraction $\Lambda\bar{e}$, where Λ denotes the “unit cost of effort”.

The critical feature of this model is the multiplicative functional form in equation (1). Shirking has a percentage effect on the CEO’s overall utility, and thus the dollar amount the CEO would pay to consume private benefits is increasing in his wage. Private benefits are therefore a normal good in our model, which is consistent with the treatment of most goods and services in consumer theory. Under this assumption, 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. As the CEO becomes richer, he purchases greater amounts of all goods and services, including private benefits. This assumption is also plausible under the literal interpretation of shirking as leisure time. Leisure allows the CEO to enjoy goods and services that he can purchase with his salary, and so shirking and consumption are complementary goods.

In addition to being intuitively appealing, multiplicative preferences also have empirically consistent implications for the scaling of labor supply with the wage. With an additive specification, leisure falls to zero as salary rises; here, labor supply is unaffected by wage changes

over time.³ This empirical consistency explains the common use of multiplicative preferences in macroeconomics, a field in which models are frequently calibrated to the data. Indeed, they are a necessary feature of macroeconomic models that feature rising wages but constant labor supply. For example, Cooley and Prescott (1995) write: “For the postwar period, [per capita leisure] has been approximately constant. We also know that real wages ... have increased steadily in the postwar period. Taken together, those two observations imply that the elasticity of substitution between consumption and leisure should be near unity.”

We now turn from preferences to the effort production function. The baseline firm value is S and the end-of-period stock price P_1 is given by⁴

$$P_1 = S(1 + \eta)(1 + e - \bar{e}), \quad (2)$$

where η is stochastic noise with mean zero. Effort has a multiplicative effect on firm value: low effort ($e = 0$) reduces firm value by a fraction \bar{e} . This is plausible for the majority of CEO actions which can be “rolled out” across the whole company, and thus have a greater effect in a larger firm. Examples include the choice of strategy, the launch of new projects, or designing a process innovation to increase production efficiency. However, certain actions have a fixed dollar effect independent of firm size, such as perk consumption or stealing. We consider such additive actions in Section 3.1.

We observe that, on the equilibrium path implementing high effort, the initial stock price is $P_0 = E[P_1]$, i.e. $P_0 = S$. We assume that $S > w\Lambda$, where S is the firm’s market capitalization: the firm value gains from high effort exceed the manager’s disutility, and so it is optimal to elicit effort.⁵ 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.

The CEO’s compensation c is composed of a fixed cash salary $f \geq 0$, and $\nu \geq 0$ shares:

$$c = f + \nu P_1. \quad (3)$$

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.

The optimal contract elicits high effort ($e = \bar{e}$) and pays the CEO his reservation wage,

³For example, consider the labor supply l of a worker living for one period, with a wage w , consumption $c = wl$, and utility $v(c, l)$. He solves $\max_l v(wl, l)$. If utility is $v(c, l) = \phi(cg(l))$, then the problem is $\max_l \phi(wl g(l))$, and the optimal labor supply l is independent of w .

⁴Here we normalize the normal growth of earnings to 0. We could normalize to another value g , by formulating the end-of-period-price as $P_1 = S(1 + \eta)(1 + e - \bar{e})(1 + g)$. The rest of the analysis would be unchanged.

⁵The proof is as follows. If the manager works, he is paid w and firm value (net of wages) is $S - w$, leading to total surplus of S . If the manager shirks, he is paid $w(1 - \Lambda\bar{e})$, to keep his utility at w . Firm value, net of wage, is $V = S(1 - \bar{e}) - w(1 - \Lambda\bar{e})$ and total surplus is $V + w = S(1 - \bar{e}) + w\Lambda\bar{e}$. Hence total surplus is higher if the manager works if and only if $S \geq S(1 - \bar{e}) + w\Lambda\bar{e}$, i.e., $S \geq w\Lambda$.

i.e. $E[c] = w$, while minimizing the number of shares given to the manager. It is stated in Proposition 1 below.

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. Specifically, it comprises a fixed base salary, f^* , and $\nu^* P_0$ worth of shares, with:*

$$\nu^* P_0 = w\Lambda, \tag{4}$$

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

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

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

where $r = P_1/P_0 - 1$ is the firm's stock market return.

In the optimal contract, $\Lambda\%$ of the CEO's wage w is paid in shares, and the remainder $(1 - \Lambda)\%$ in cash. The intuition follows from our multiplicative specification. The utility gained from shirking is increasing in the CEO's dollar wage. The cost of shirking has a percentage effect on firm value, and thus is increasing in the dollar value of the CEO's shares. Hence, to maintain equality between costs and benefits, any increase in the CEO's wage must be matched by an exactly proportional increase in his shares – in other words, the CEO's dollar equity must comprise a constant fraction of the total wage.

Put differently, if effort has multiplicative costs and benefits, the percentage change in pay for a percentage change in firm value (i.e. “%-%” incentives) is the relevant measure, and must be at least Λ to achieve incentive capacity. In terms of real variables, this “%-%” measure equals the proportion of total salary that is comprised of shares. The optimal contract minimizes the number of shares, and so this proportion equals Λ .

Note that any contract where at least $\Lambda\%$ of the wage is in shares will achieve incentive compatibility, and there are a continuum of contracts that satisfy this criterion. The model's strongest prediction is thus in the form of an inequality restriction; the optimal ratio is not steadfastly determined. We choose the contract that minimizes the number of shares as this would be strictly optimal under any non-zero level of risk aversion. However, if risk considerations are insignificant in reality, the ratio may exceed Λ in some cases, and the model's empirical implications will be contradicted. We show in Section 2 that its main predictions are quantitatively consistent with the data.

1.2 Incentive Pay in Market Equilibrium

The above principal-agent model only solves for the optimal division of a fixed wage w into cash and shares. We now embed the previous analysis into a market equilibrium where the equilibrium wage w is endogenously determined. We directly import the model of Gabaix and Landier (2008) (“GL”), the essentials of which we review in Appendix A. There is a continuum of firms of different size and managers with different talent. Since talented CEOs are more valuable in larger firms, the n th most talented manager is matched with the n th largest firm in competitive equilibrium, and earns the following competitive equilibrium pay:

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

where $S(n)$ is the size of firm n , n_* is the index of a reference firm (e.g. the 250-th largest firm), $S(n_*)$ is the size of that reference firm, $D(n_*)$ is a constant independent of firm size, and α , β and γ are also constants. In particular, CEOs at large firms earn more as they are the most talented, with a pay-firm size elasticity of $\rho = \gamma - \beta/\alpha$. For their calibration, GL use $\alpha = \gamma = 1$, $\beta = 2/3$.

In our model, firm values P_0 and P_1 are endogenous to CEO effort, but baseline firm size S is exogenous. The incentive problem is unchanged even if S is endogenous (e.g. to CEO talent). It remains the case that firm value falls by $\bar{e}\%$ if the manager shirks, and so Proposition 1 continues to hold. In addition, GL give several reasons why exogenous firm size is a reasonable benchmark for the talent assignment model (see, e.g., their footnote 11 and Online Appendix). In particular, the calibrations of CEO talent by GL and Tervio (2007) evaluate the impact of CEO talent on size to be very small. Therefore, size is primarily determined by factors other than CEO talent (such as productivity differentials as in Luttmer (2007)). Tervio (2007, Section 3.1) shows that the scaling of CEO talent impact is robust to some forms of endogenous firm size. Endogenizing firm size is the focus of Baranchuk, MacDonald and Yang (2007).

GL only specify the total compensation that the CEO must be paid in market equilibrium. We now incorporate the incentive results of Section 1.1 to determine the form of compensation. The equilibrium incentive pay is analogous to Proposition 1 and stated below:

Proposition 2 (*CEO incentive pay in market equilibrium*). *Assume $\Lambda < 1$ (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:*

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

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. The optimal contract pays manager n a fixed base salary, f_n^* , and

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

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

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

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

where $r(n) = P_{1n}/P_{0n} - 1$ is the firm's stock market return.

In Proposition 2, there is a full separation between the determination of expected pay (which is the same as in GL), and the determination of the incentive mix (which is the same as in Proposition 1). The reason is that the equilibrium entails a first-best level of effort, and all CEOs exert the same high effort. Firms therefore compete on pay, not on required effort, and so total pay is as in GL. Given this total pay, Proposition 1 an optimal way a firm incentivizes the CEO to achieve the high effort.

We assume that Λ is constant across firms to maintain the simplicity of our model and limit our degrees of freedom in calibration. The above Proposition remains valid if Λ differs across firms: Λ is simply replaced by 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, all firms are at their "baseline" value of S as in GL, and so CEO assignment is unchanged.

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 their effort. 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-effort."

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 compensation, S the market

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

$$b^I = \frac{\partial c}{\partial r} \frac{1}{w} = \frac{\Delta \ln \text{Pay}}{\Delta \ln \text{Firm Value}} \quad (9)$$

$$b^{II} = \frac{\partial c}{\partial r} \frac{1}{S} = \frac{\Delta \$\text{Pay}}{\Delta \$\text{Firm Value}} \quad (10)$$

$$b^{III} = \frac{\partial c}{\partial r} = \frac{\Delta \$\text{Pay}}{\Delta \ln \text{Firm Value}}. \quad (11)$$

b^I denotes %-%-incentives and is used (or advocated) by Murphy (1985), Gibbons and Murphy (1992) and Rosen (1992). b^{II} represents \$-\$ incentives and is used by Demsetz and Lehn (1985), Yermack (1995) and Schaefer (1998). b^{III} measures \$-\$% incentives, the dollar change in pay for a given percentage change in firm value, and is advocated by Holmstrom (1992). The next Proposition derives predictions for these quantities.

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

$$b^I = \Lambda \quad (12)$$

$$b^{II} = \Lambda \frac{w}{S} \quad (13)$$

$$b^{III} = \Lambda w, \quad (14)$$

where w is given by (7).

Share-based compensation can be implemented in a number of forms, such as stock, options and bonuses. 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, where instruments are converted into share equivalents according to their delta (see Section D for the conversion of options.)

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

1. b^I is independent of firm size:

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

2. b^{II} scales as $S^{\rho-1}$:

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

3. b^{III} scales as S^ρ :

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

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

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

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 $S(n_*)$ as follows:

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

where γ is the elasticity of CEO impact in GL (see equation (29) in Appendix A

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

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

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

Insert Table 1 about here

Propositions 4 and 5 imply that the %-% measure of pay-performance sensitivity is independent of both firm size and the size of reference firms. From Proposition 1, %-% incentives equal Λ in the optimal contract. Since Λ is constant across firms, %-% incentives should also be constant if compensation is efficient in all firms.

In an additive model, effort has a fixed dollar effect on firm value and the manager's utility. Thus, \$-\$ incentives (b^{II}) are the relevant measure and should be constant across firms, if all companies are contracting optimally. However, Demsetz and Lehn (1985), Jensen and Murphy (1990), Gibbons and Murphy (1992), Schaefer (1998), Hall and Liebman (1998) and Baker and Hall (2004) all find that \$-\$ incentives decline strongly with firm size. One common interpretation of this result is that incentives are suboptimally low in large firms, either because of greater managerial entrenchment in such companies (Bebchuk and Fried (2004)), or because large firms are highly visible and face strong political constraints on high pay (Jensen and Murphy (1990)).

However, Proposition 4 has a different conclusion: b^{II} should *optimally* decline with firm size. 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 have low \$-\$ incentives. This point has been previously noted by Hall and Liebman (1998) and modeled by Baker and Hall (2004) in a different framework, to back out the production function that would be consistent with observed incentives. We postulate multiplicative specifications based on first principles and derive quantitative predictions for this scaling in market equilibrium. Since $b^{II} = b^I \frac{w}{S}$ and the wage w scales with $S^{1/3}$, b^{II} should scale with $S^{-2/3}$.

Finally, Section 1.1 shows that b^{III} 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.

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. Since effort continues to have a multiplicative impact on firm value and utility, it remains the case that %-%- incentives should be independent of firm size.

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, 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:*

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

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

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

B^{II} is used by Jensen and Murphy (1990); Hall and Liebman (1998) report both B^{II} and B^{III} .⁶ Multiplying the pay-performance sensitivities in Proposition 5 by $\frac{W}{w}$ gives the following magnitudes for wealth-performance sensitivities:

⁶We 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.

Proposition 6 (*Wealth-performance sensitivities*). *Let W denote total CEO wealth 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, but stems from the assumption that shirking is a normal good.

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 described later in Section 2, we confirm that this elasticity holds for the relationship between wealth and size: we find a coefficient of 0.37 with a standard error of 0.05. By contrast, W/w has an elasticity of 0.02 (standard error of 0.07). 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.

1.5 The Requirement for Multiplicative Preferences

Our choice of the multiplicative specification (1) was motivated by first principles, in particular the view that private benefits are most plausibly a normal good. Such a functional form led to the prediction that B^I is independent of w , which we validate empirically in Section 2.1. We now demonstrate that additive preferences would achieve different predictions; indeed, we prove that multiplicative preferences are necessary (as well as merely sufficient) to yield this implication. For clarity, we use a one-period model and focus on the analogous measure b^I .

Many previous theories of CEO pay (Haubrich (1994), Schaefer (1998), Baker and Hall (2004)) are based on the classical “additive” models of Grossman and Hart (1983) and Holmstrom and Milgrom (1987), which in its risk-neutral version uses the form $E[c] - h(e)$, with h non-decreasing. We maintain the same contract structure (equation (3)): b^I is the fraction of w invested in stock, so that $c = w(1 + b^I r)$. With the utility function $E[c] - h(e)$, the optimal

b^I is given by $b^I = \frac{h(\bar{e})-h(0)}{w\bar{e}}$, which implies:⁷

$$b^I \propto w^{-1}. \quad (22)$$

The additive form therefore predicts that b^I decreases with the wage, whereas the multiplicative model predicts that b^I is constant.

Another popular utility function is $E[c^\Gamma/\alpha] - h(e)$, with $\Gamma \in (0, 1]$. This leads to $b^I \propto w^{-\Gamma}$ for large w , and thus also predicts that b^I declines with firm size. For sufficiently high consumption, effort has a very small negative 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 . We consider a general utility function $E[u(c, e)]$, with $e \in \{0, \bar{e}\}$. The firm's return is $r = e - \bar{e}$, so that the return is 0 on the equilibrium path where the CEO exerts high effort. The firm selects expected pay w and incentives b^I so that $c = w(1 + b^I r)$. The optimal contract minimizes w and b^I while granting the CEO his reservation utility of \underline{u} and eliciting $e = \bar{e}$.⁸ The next Proposition states that multiplicative preferences are required for the optimal b^I to be independent of \underline{u} (and thus w).

Proposition 7 (*Necessity and sufficiency of multiplicative preferences to generate a size-independent b^I*). *Assume the CEO's utility function is $u(c, e)$, and the firm's return is $r = e - \bar{e}$. Suppose the optimal affine contract involves a scaled pay-performance sensitivity $b^I = E[\partial c / \partial r] / E[c]$ that is independent of $E[c]$. Then, the utility function is multiplicative in consumption and effort, i.e. can be written:*

$$u(c, e) = \phi(cg(e)) \quad (23)$$

for some functions ϕ and g .

Conversely, if preferences are of the type (23), then the optimal contract has a slope b^I that is independent of $E[c]$.

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.

To keep the analysis streamlined, we proved the above Proposition in a restrictive context with no noise, although we allowed for a general utility function. We suspect that the results

⁷The proof is as follows. The optimal b^I is the smallest b^I such that $E[c - h(\bar{e}) | e = \bar{e}] \geq E[c - h(0) | e = 0]$, and so satisfies $E[c - h(\bar{e}) | e = \bar{e}] = E[c - h(0) | e = 0]$. Since $c = w(1 + b^I r)$, $E[c | e] = w(1 + b^I(e - \bar{e}))$. Therefore, $w - h(\bar{e}) = w(1 - b^I \bar{e}) - h(0)$, i.e. $b^I = \frac{h(\bar{e})-h(0)}{w\bar{e}}$.

⁸More fully, $\underline{u} = E[u(c, e) | e = \bar{e}] \geq E[u(c, e) | e = 0]$.

extend to more general settings, but such an investigation is beyond the central objective of this paper.⁹

2 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 2.1 shows that the data quantitatively matches the model’s predictions for the scalings of incentives with firm size. Section 2.2 demonstrates that the level of incentives is also consistent with our optimal contracting model.

2.1 Determinants of CEO Incentives

As noted in Section 1.3, the stylized fact is that \$-\$ incentives decline with firm size. Our market equilibrium model derives a quantitative prediction for this scaling. Specifically, $\gamma - \beta/\alpha = 1/3$ (as found by GL) implies an elasticity of $-2/3$. Consistent with our model, Schaefer finds \$-\$ incentives scale as $S^{-\xi}$, with $\xi \simeq 0.68$.¹⁰ Existing research is also consistent with the model’s prediction that %-% incentives are independent of size (Gibbons and Murphy (1992)). We do not know of any studies that investigate the link between \$-% incentives and size.

However, prior findings cannot be interpreted as conclusive support of the model. The vast majority of CEO incentives come from his existing *stock* of shares and options, rather than compensation *flows* (salary, bonus and new grants of stock and options). Owing to data limitations, Gibbons and Murphy (1992) consider only flow compensation, and Schaefer (1998) includes existing stock, but not options. We therefore conduct our own empirical tests of the model, using measures of wealth-performance sensitivity. We merge Compustat with ExecuComp (1992-2006) and select the largest 500 firms in aggregate value (debt plus equity)

⁹For instance, with noise, we suspect that to keep b constant across expected utilities, the function ϕ must actually be: $\phi(c) = A \ln c + B$ or $Ac^{1-\Gamma}/(1-\Gamma) + B$.

¹⁰This 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.

in each year.¹¹ We calculate the wealth-performance sensitivities as follows¹²:

$$B^I = \frac{1}{w_t} \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P \right] \quad (24)$$

$$B^{II} = \frac{1}{S_t} \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P \right] \quad (25)$$

$$B^{III} = \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P \right] \quad (26)$$

We use the Core and Guay (2002) methodology to estimate the option deltas. (Appendix D describes our calculations in further detail.) All variables are converted into constant dollars using the GDP deflator from the Bureau of Economic Analysis. Controlling for year and Fama-French (1997) industry fixed effects, and clustering standard errors at the firm level, we estimate the following elasticities:¹³

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

where S is the firm's aggregate value of debt plus equity. Table 2 illustrates the results, which are consistent with the predictions of equation (15).¹⁴ Specifically, B^I is independent of firm size: the coefficient of 0.04 is less than its standard deviation. B^{II} (B^{III}) have size elasticities of -0.60 (0.40), 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.

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 relevant measure depends on the scaling of

¹¹Our results are very similar if we use sales as a measure of firm size, and if we select the top 1000 or 200 firms.

¹² $B^{III} = \frac{\partial W}{\partial r} = \frac{\partial W}{\partial S} P$, where $\frac{\partial W}{\partial S}$ is the "delta" of the CEO's portfolio. The delta of each share is 1, and so the delta of his stock holdings equals the number of his shares. The delta of each option is $\frac{\partial V}{\partial P}$ and so the delta of his option holdings equals $\frac{\partial V}{\partial P}$ multiplied by the number of options. Multiplying both components by P gives $\frac{\partial W}{\partial r}$, i.e. B^{III} . B^I and B^{II} are transformations of B^{III} as given by equations (19) and (20).

¹³We 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.

¹⁴Although we have 15 years of data and 500 firms, there are fewer than 7,500 observations in each regression, mainly because a number of firms do not have SIC codes and thus cannot be classified into an industry.

CEO productivity with firm size. If productivity is constant in dollar terms regardless of firm size, B^{II} is appropriate as it is size-invariant; if it is linear in firm size, 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.

We show that the optimal incentives measure depends on the specification for preferences as well as the production function. In our model, utility is multiplicative in effort and we predict that B^I is independent of firm size. Table 2 empirically confirms the size invariance of B^I , thus supporting our modeling assumptions, as well as the size dependence of B^{II} and B^{III} . We thus advocate B^I as the preferred empirical measure of incentives. If incentives are the dependent variable, size independence allows comparability of the strength of incentives across firms or over time. If wealth-performance sensitivity are the independent variable of interest, size invariance ensures that its explanatory power does not simply arise because it proxies for size. If size (or a variable correlated with size) is the covariate of interest and incentives are merely a control, the use of B^I ensures that the coefficient on size is not distorted by the inclusion of another size proxy in the regression.

To our knowledge, B^I has not been used in previous studies. Murphy (1985) and Gibbons and Murphy (1992) calculate the elasticity of flow pay to firm value, i.e. b^I . Hall and Liebman (1998) calculate 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. In addition to introducing B^I empirically, we justify it theoretically by comparing its properties to alternative measures. Our theoretical framework also underpins our definition of B^I , in particular why flow compensation only should be in the denominator.

In a contemporaneous paper, Edmans, Goldstein and Jiang (2008) use stochastic frontier analysis to construct a measure of each firm's maximum potential valuation under full efficiency. They find that a firm's discount to its potential value is strongly decreasing in B^I . By contrast, Habib and Ljungqvist's (2005) stochastic frontier analysis finds no relationship with B^{II} . This supports the view that B^I is a well-behaved measure of incentives.

2.2 The Level of CEO Incentives

Having investigated our model's scaling predictions, we now assess whether empirical levels of wealth-performance sensitivity are consistent with efficiency. Our primary measure is %-% incentives; the other measures are mechanical transformations. The model predicts $B^I = \Lambda \frac{W}{w}$ (equation (19)). We present figures for 1999, the median year in our sample by level of incentives. The median B^I in 1999 is 9.05.¹⁵

¹⁵Hall and Liebman (1998, Table VIII) estimate $B^I = 3.9$ for 1994, the final year in their sample. Their denominator includes not only flow compensation but also the expected appreciation of the CEO's stock and options.

We therefore calibrate $\Lambda = B^I w/W = 9w/W$. Shirking increases the CEO’s utility by a fraction $\Lambda \bar{e} = 9\frac{w}{W}\bar{e}$ of his wealth, i.e. $\$9w\bar{e}$ in dollar terms. \bar{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%.¹⁶ However, the takeover premium can be motivated by factors other than managerial misbehavior, such as synergies or undervaluation. Since a high input for \bar{e} would make it easier to match the B^I found in the data, we conservatively set $\bar{e} \simeq 10\%$ which yields $\Lambda \bar{e} W = \$0.9w$. The current level of incentives is able to deter multiplicative 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 seems that observed incentives are able to address a number of agency issues with multiplicative effects on firm value. This result echoes Taylor (2007), who builds and estimates a structural model and finds that the observed level of CEO turnover is not too low to be consistent with optimal firing decisions. However, incentives are not effective in two cases: if the utility from shirking is very high (Proposition 1 requires $\Lambda < 1$ for shirking to be preventable), 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: if the manager fails to resign when it is optimal, he retains his salary (plus private benefits of control) in many future years, the present value of which may plausibly exceed his annual pay. Another is expansive acquisitions, since Bebchuk and Grinstein (2007) find that increases in firm size lead to higher CEO pay in future periods. Moreover, our estimate of $\$0.9w$ hinges upon our chosen input for \bar{e} (it does not require an estimation of W/w , since this cancels out). For actions with smaller effects on the stock price, observed incentives will be too low to deter misbehavior. In particular, in Section 3.1 we show that actions with additive effects on firm value have a small impact on equity returns in large firms, and cannot be deterred by incentives.

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.¹⁷ However, ExecuComp provides data on a CEO’s financial wealth in his own firm. In 1999, we estimate a median value of Financial wealth in the firm / Pay equal to 9.44. We assume that the CEO’s wealth in his own firm is half his total financial wealth, and that his human wealth (NPV of future wages) approximately equals his entire financial wealth. This leads to an estimate of W/w of 37.8. We therefore have $\Lambda \bar{e} = 9\frac{w}{W}0.1 \simeq \frac{0.9}{37.8} = 0.024$. This means that, if the CEO shirks, his utility increases by an amount equivalent to 2.4% of his wealth.

Since B^{II} and B^{III} are mathematically linked to B^I , our ability to explain B^I means

¹⁶Bennedsen, 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.

¹⁷We thank David Yermack for discussions on this point. See Becker (2006) for a study with Swedish CEOs.

that the model can also match these other measures of wealth-performance sensitivity. For example, $B^{II} = B^I \frac{w}{S}$. The median size of the top 500 firms in 1999 is \$19 billion, with median pay of \$6.2 million. $B^I = 9$ is therefore consistent with a Jensen-Murphy semi-elasticity of $B^{II} = 9 \times (\$6.2 \text{ million}) / (\$19 \text{ billion})$. This represents a wealth rise of \$2.94 for a \$1,000 increase in firm value, close to our directly measured figure of \$2.62.¹⁸

3 Extensions

This section considers extensions and other specifications of the model. Section 3.1 shows that incentives are ineffective for deterring actions that are additive in firm value, such as perk consumption. Section 3.2 extends the model to multiple effort levels and derives a positive relationship between corporate governance and incentives. In Section 3.3 we show that our model predicts a positive association between firm volatility and wealth volatility, which we support empirically. By contrast, traditional frameworks have the opposite prediction. Section 3.4 shows how our model explains the empirical results of Baker and Hall (2004).

3.1 Additive Production Functions and Perks

In the core model, effort has a multiplicative effect on firm value. This allows all incentive problems to be solvable through the contract specified in Proposition 1 (as long as $\Lambda < 1$). Since the majority of CEO actions (e.g. strategy choice or process innovation) can be “rolled out” across the entire firm, the multiplicative specification likely holds for many managerial decisions. However, perk consumption in particular is likely to have an additive effect on firm value. For example, purchasing an unnecessary corporate jet for a dollar value \bar{L} , or stealing \bar{L} , reduces firm value by \bar{L} regardless of firm size. The following Proposition states that incentives are unable to deter such actions.

Proposition 8 (*Impossibility of deterring perk consumption through incentive pay*). *Assume that effort $e = \underline{e}$ (i.e. perk consumption) reduces firm value by \bar{L} dollars. Assume that $\bar{L} > w\Lambda\bar{e}$, so that perk prevention would maximize total surplus. It is impossible to prevent perk consumption if $S > \bar{L}/(\Lambda\bar{e})$, i.e. the firm is sufficiently large.*

Hence if $w\Lambda\bar{e} < \bar{L} < S\Lambda\bar{e}$, perk consumption is inefficient but cannot be prevented with shares.¹⁹ Since the perk is fixed in absolute terms, the stock price of a large firm is relatively

¹⁸\$2.94 is different from the directly-measured number of \$2.62, as the median size firm does not have the median level of incentives. \$2.94 is smaller than the \$5.29 reported by Hall and Liebman (1998) for 1994, their final year, and the \$3.25 reported by Jensen and Murphy (1990) because we are considering only the top 500 firms and B^{II} declines with size. Across the whole sample, the median is \$8.74.

¹⁹Edmans and Gabaix (2008) extend the model to incorporate general incentive contracts and risk aversion. Perks can be prevented with highly nonlinear contracts, but these impose such high risk on the CEO that total surplus falls with perk prevention. Thus, it remains the case that incentives are ineffective at deterring perks.

insensitive to perk consumption: stock returns only fall by \bar{L}/S . 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. salary 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 $w\Lambda\bar{e}$) so incentive compatibility is still not achieved. Another 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 $f \geq 0$ owing to limited liability.

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 strategy choice. However, perks are typically independent of firm value, and thus cannot be addressed by incentives. As with multiplicative actions where Λ is high, perks should instead be controlled by active corporate governance such as direct monitoring. 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)) 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.

Effective monitoring, however, may be difficult to achieve, particularly since governance may be endogenously chosen by the CEO (e.g. Hermalin and Weisbach (1998)) and so perks are often consumed in reality (Yermack (2006)). Moreover, governance is primarily effective at punishing errors of commission (reducing firm value) rather than errors of omission (failing to take an action that increases firm value). This is because the board is highly unlikely to know the set of value-enhancing actions the manager can undertake: it cannot punish a CEO for failing to invent a new product, since it would not have the idea that such a product could be created. Hence, active monitoring and incentives should be used in tandem: the former to deter additive value-destructive actions, and the latter to encourage multiplicative value-enhancing effort.

Overall, 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 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 for shareholders 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.

3.2 Poor Corporate Governance and Incentives

This section extends the model to a continuum of effort levels. This analysis shows that our results are not dependent on the binary specification of effort that we used for tractability. Moreover, it allows us to examine the effect of corporate governance on incentives. As before, the maximum effort level is optimal. While there are many possible ways to model poor corporate governance, we represent it as the board setting a target effort level below the maximum.²⁰

In the extended model, the CEO can choose an effort level $e \in [\underline{e}, \bar{e}]$. The CEO's utility function is $E[cg(e)]$, where $g(e)$ is decreasing and $\ln g(e)$ is concave; the latter is a standard assumption to ensure that the utility function is log concave. The board sets a target level $\hat{e} < \bar{e}$. The next Proposition derives the corresponding incentive level.

Proposition 9 (*Negative relationship between governance and incentives*). *Suppose that the board wishes to implement an effort level $\hat{e} \in (\underline{e}, \bar{e})$. Then board sets an incentive level of:*

$$b^I(\hat{e}) = \frac{-g'(\hat{e})}{g(\hat{e})} (1 + \hat{e} - \bar{e}) > 0 \quad (27)$$

%-% incentives $b^I(\hat{e})$ are increasing in \hat{e} . The contract comprises a fixed base salary of $f^* = w(1 - b^I(\hat{e}))$ and $\nu^* P_0 = b^I(\hat{e})w$ worth of shares. To implement $\hat{e} = \bar{e}$, the board must set $b^I \geq \frac{-g'(\bar{e})}{g(\bar{e})}$.

Thus a poorly governed firm will set a lower level of incentives, in turn allowing shirking. To evaluate this prediction empirically, we proceed as in Table 2 and add the Gompers, Ishii and Metrick (2003) governance index as an additional explanatory variable in the regression of B^I on firm size. We find a coefficient of -0.048 , with a t -statistic of -2.22 , which supports Proposition 9. The size elasticity becomes 0.01 , even closer to the predicted value of 0 . 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 13% .

The relationship between governance and incentives may explain the rise in wealth-performance sensitivity over time (documented by Hall and Liebman (1998), Murphy (1999) and Frydman and Saks (2007)). Corporate governance has likely strengthened in recent years from changes resulting from recommendations and legislation (such as the 1992 Cadbury Report and the 2002 Sarbanes-Oxley Act), changes enforced by activist shareholders (see, e.g., Carleton, Nelson and

²⁰Note that allowing shirking is a costly way to favor the CEO, since shirking has a multiplicative effect on firm value. A more efficient method would be to maintain optimal incentives, but to give the manager superfluous cash.

Weisbach (1998)) or voluntary changes resulting from increased investor and media scrutiny of governance, such as the removal of board interlocks. Improvements in corporate governance will lead to a rise in \hat{e} and thus an increase in incentives. In addition, deregulation and globalization have plausibly increased the manager’s scope for creating value. This augments \bar{e} and thus the optimal wealth-performance sensitivity.²¹

3.3 The Link Between CEO Wealth Volatility and Firm Volatility

This section contrasts the opposite predictions of our model and standard models for the relationship between wealth volatility and firm volatility. Our model predicts a positive association, which we support empirically. However, standard models feature a trade-off between incentives and risk, thus implying a negative correlation.

We first review the “trade-off” prediction of standard models. One variant features additive preferences and a multiplicative production function, but we later also consider additive production functions. To simplify the exposition, we use the certainty-equivalent representation of the 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 $S_1 = S(1 + Le + \eta)$, where L measures the productivity of effort and η is normal noise with mean 0 and variance σ_r^2 . The firm maximizes $S(1 + Le) - E[c]$, its expected value next period net of CEO pay. As before, compensation comprises fixed pay f , plus ν shares.

Under this model, the optimal \$-\$ incentives are $b^{II} = \partial c / \partial S_1 = L / (L^2 + a\sigma_r^2)$, and thus decreasing in firm volatility.²² This arises because 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, incentives impose even higher costs on the manager, and thus the optimal level is lower.

In addition to predicting a negative relationship between b^{II} and σ_r , standard models also 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 $\partial stdev(c) / \partial \sigma_r = -S(1 - 2b^{II})b^{II}$. 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 CEO wealth volatility declines in firm volatility.²³

²¹Note that the rise in incentives may also be for reasons outside the model. For example, until recently, at-the-money options did not need to be expensed and thus may have been used as “hidden” compensation. Alternatively, they may have been a mechanism to avoid the additional tax liability caused by granting a cash salary in excess of \$1 million.

²²Normalizing the initial share price to $P_0 = 1$, 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, $S(1 + \nu L^2) - \frac{a}{2}\sigma_r^2\nu^2 - \frac{\nu^2 L^2}{2}$, and selects $\nu = SL^2 / (L^2 + a\sigma_r^2)$. The CEO’s total pay is therefore $c = f + S_1 L / (L^2 + a\sigma_r^2)$.

²³The standard 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

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.

$$stdev(W_{t+1} - W_t) = B^{III}\sigma_r \propto S^\rho\sigma_r, \quad (28)$$

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 evaluate these contrasting predictions empirically. As discussed more fully in Appendix D, 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 2, i.e. $stdev(W_{t+1} - W_t) = B_t^{III}\sigma_r$.²⁴ The second uses ex post realized volatility, i.e. $stdev(W_{t+1} - W_t) = \ln|W_{t+1} - W_t|$. 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 $stdev(W_{t+1} - W_t) = \beta_S \ln S + \beta_\sigma \ln \sigma_r$ 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 increasing in 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.

We now detail the origins of the contrasting predictions. In our model, incentives ensure maximum effort regardless of the cost imposed on the manager. 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 incentives)²⁵, it is always efficient to implement the maximum level of effort. Simply put, risk does not affect incentives because it is second-order.

The absence of a trade-off results from two features of our model: the existence of a maximum effort level, and a multiplicative production function which means that maximum effort is optimal. A maximum will exist either because there is a limit to the number of productive

linear in wealth in the standard model.

²⁴Indeed, for small time intervals, $W_{t+1} - W_t = W_t'(r) r_t = B^{III}r_t$, so $stdev(W_{t+1} - W_t) = B^{III}stdev(r_t) = B^{III}\sigma_r$.

²⁵In this paper, the only cost of incentives is the direct disutility from working. In Edmans and Gabaix (2008) we show that introducing risk aversion does not change this result. The intuition is as in the core model: the cost of risk-bearing is a function of the manager, and thus much smaller than the benefits of effort.

activities that a CEO can undertake (e.g. finite NPV-positive projects) or a limit to the number of hours in a day the CEO can optimally work. Models with binary effort levels also assume a maximum; our model is more general as it allows for intermediate effort levels (see Section 3.2). Introducing a maximum \bar{e} into the standard model with a multiplicative production function and additive preferences would also generate a corner solution if $SL^3/(L^2 + a\sigma_r^2) > \bar{e}$. With an additive production function ($S_1 = S + Le + \eta$), the required condition is $L^3/(L^2 + a\sigma_r^2 S^2) > \bar{e}$ and is less likely to be satisfied for large firms. Hence, the combination of a maximum effort level and a multiplicative production function is necessary to remove any trade-offs.

Consistent with our evidence, Prendergast's (2002) survey finds that most studies detect no relationship between incentives and firm risk, with the remainder equally divided between reporting positive and negative associations. He models an explanation based on the allocation of responsibility to employees; ours is based on the observation that the cost of risk is very small relative to the firm, so that trade-off considerations are insignificant. Our model provides another explanation for Prendergast's puzzle.

3.4 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 additive preferences, which requires L (the productivity of effort) to be size-dependent for B^{II} to decline with size. They therefore use their results to back out the required scaling of L with size. In our model, preferences and production functions are motivated by first principles. We demonstrate that these specifications can generate the empirical scalings.

Baker and Hall estimate a functional form for $L(e, S)$. They derive an equation for I , the CEO's dollar productivity, as a function of firm size; in our notation, $I = LS$. Their equation (3) predicts $I^{BH} = \sqrt{\frac{2b^{II}a}{1-b^{II}}}\sigma_r S$, where a is the coefficient of absolute risk aversion. 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 their specification (2), they assume wealth is proportional to the CEO's wealth invested in the firm, and so $a \propto W^{-1}$. Since $w \propto W$ empirically (see section 1.4), specifications (1) and (2) both lead to $a \propto w^{-1}$.

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),²⁶ 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.

²⁶Regressing log volatility on log aggregate value, year dummies and industry dummies yields an insignificant coefficient of -0.0019 (standard error of 0.0124).

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 empirical results can also be quantitatively explained by our model.

4 Conclusion

This model studies optimal executive compensation in a setting with two unique features. First, motivated by first principles, we depart from traditional additive specifications and assume that effort has a multiplicative effect on both firm value and CEO utility. Second, while principal-agent models are typically partial equilibrium and focus on the composition of a fixed level of total pay, we endogenize salary by embedding the agency problem in a competitive assignment model. The unified framework has a number of empirical implications which differ from standard models with linear functional forms:

(i) Dollar-dollar incentives 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-percent incentives should have a size elasticity of $1/3$.

(ii) Scaled wealth-performance sensitivity (percent-percent incentives, i.e. the dollar change in wealth for a percentage change in firm value, scaled by annual pay) is invariant to firm size.

(iii) Observed levels of wealth-performance sensitivity (percent-percent incentives) are sufficient to deter value-destructive actions that yield private benefits no greater than 0.9 times the annual wage. Similarly, the level of dollar-dollar incentives should be very small, as empirically documented by Jensen-Murphy (1990).

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

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

While our model shows that a number of observed features of compensation are consistent with an optimal contracting model, there are number of stylized facts upon which the model is silent, such as the use of pensions and other forms of "hidden" compensation, the use of accounting-based compensation (such as bonuses) and severance pay. Equally, there are many other determinants of compensation that the model does not consider, such as risk, stockholder-bondholder conflicts, screening of ability, managerial entrenchment, and the inducement of innovation. Owing to its tractability and empirical consistency, our model may provide a useful

benchmark on which future models can be built to explore these additional determinants. The resulting empirical predictions can be compared with the above stylized facts to assess whether they are consistent with optimal contracting or instead result from rent extraction. 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?²⁷ How much of the time series variation in incentives, documented by Frydman and Saks (2007), can be explained by our model?

One important caveat is that we used the empirical consistency of our model’s predictions to justify our assumed functional form, 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. 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 On the equilibrium path (where he exerts a high effort \bar{e}), the manager should earn his market wage: $E[c | e = \bar{e}] = w$. Using $c = f + vP_1$, we calculate:

$$\begin{aligned} E[c | e = \bar{e}] &= f + \nu E[P_1] = f + \nu P_0 = w \\ E[c | e = 0] &= f + \nu E[P_1] (1 - \bar{e}) = f + \nu P_0 (1 - \bar{e}) = f + \nu P_0 - \nu P_0 \bar{e} = w - \nu P_0 \bar{e}. \end{aligned}$$

²⁷Given 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).

The manager chooses $e = \bar{e}$ if:

$$E [cg(\bar{e}) | e = \bar{e}] \geq E [cg(0) | e = 0].$$

Since $g(\bar{e}) = 1$ and $g(0) = \frac{1}{1-\Lambda\bar{e}}$, this incentive-compatibility constraint writes:

$$w \geq \frac{w - \nu P_0 \bar{e}}{1 - \Lambda \bar{e}} \Leftrightarrow \nu P_0 \geq w \Lambda = \nu^* P_0.$$

f^* is chosen to ensure that expected pay is w : $f^* = w - \nu^* P_0 = w(1 - \Lambda)$.

Proof of Proposition 2 Once a CEO is assigned to a firm, they reach the efficient outcome, which is to implement the high level of effort. Hence, the incentive scheme of Proposition 1 applies.

To endogenize the wage, 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

$$S' = S + CTS^\gamma, \tag{29}$$

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)):

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

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 $n/N \rightarrow 0$:

$$w(n) = \int_n^N A^\gamma BC u^{-\alpha\gamma+\beta-1} du + \underline{w} = \frac{A^\gamma BC}{\alpha\gamma - \beta} [n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}] + \underline{w}_N \sim \frac{A^\gamma BC}{\alpha\gamma - \beta} n^{-(\alpha\gamma-\beta)}. \quad (31)$$

To interpret equation (31), 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.²⁸

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.

Proof of Proposition 7 Define $\phi(c) = u(c, \bar{e})$, $g(\bar{e}) = 1$ and $g(0) = 1/(1 - b^I \bar{e})$. Since b^I achieves the minimum slope while maintaining incentive compatibility, $E[u(c, e) | e = 0] = E[u(c, e) | e = \bar{e}]$, i.e.

$$u(w(1 - b^I \bar{e}), 0) = u(w, \bar{e}) = \phi(w)$$

and so $u(c, 0) = \phi(c/(1 - b^I \bar{e})) = \phi(cg(0))$. Therefore, $u(c, e) = \phi(cg(e))$ for all c and $e \in \{0, \bar{e}\}$.

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

²⁸The derivation is as follows. Since $S = An^{-\alpha}$, $S(n_*) = An_*^{-\alpha}$, $n_* T'(n_*) = -Bn_*^\beta$, we can rewrite equation (31) as follows:

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

Proof of Proposition 8 If perk consumption occurs, $E[P_1] = S - \bar{L}$, else $E[P_1] = S$. To deter perk consumption, we require that the CEO's utility is greater under high effort:

$$f + \nu S \geq \frac{f + \nu(S - \bar{L})}{1 - \Lambda \bar{e}}.$$

Therefore $\nu(\bar{L} - S\Lambda\bar{e}) \geq f\Lambda\bar{e}$. Since $f \geq 0$ and $\nu \geq 0$, this cannot be satisfied if $S > \bar{L}/(\Lambda\bar{e})$. Finally, perk consumption is inefficient if and only if $\bar{L} > w\Lambda\bar{e}$, by the same reasoning as in footnote 5.

Proof of Proposition 9 As $P_1 = S(1 + \eta)(1 + e - \bar{e})$, and the market correctly anticipates that effort level \hat{e} is implemented, the initial price is $P_0 = E[P_1 | e = \hat{e}] = S(1 + e - \hat{e})$. The firm return is $r = P_1/P_0 - 1$. Hence, if the CEO exerts effort e , the expected return is:

$$E[r | e] = \frac{e - \hat{e}}{1 + \hat{e} - \bar{e}} \quad (32)$$

Suppose that the CEO receives νP_0 in shares, $w - \nu P_0$ is cash. His consumption is:

$$c = w - \nu P_0 + \nu P_1 = w - \nu P_0 + \nu P_0(1 + r) = w \left(1 + \frac{\nu P_0}{w} r\right) = w(1 + b^I r).$$

As before, $b^I = \frac{\nu P_0}{w}$. Hence the CEO's expected utility is:

$$E[cg(e) | e] = wg(e) \left(1 + b^I \frac{e - \hat{e}}{1 + \hat{e} - \bar{e}}\right).$$

The CEO chooses the effort $\hat{e} = \arg \max_e E[cg(e) | e]$ if and only if:

$$wg'(\hat{e}) + wg(\hat{e}) b^I \frac{1}{1 + \hat{e} - \bar{e}} = 0.$$

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. It shows that the key results of the one-period model still hold: since effort continues to have multiplicative costs and benefits, %-%-incentives remain relevant.

As in the core model, we retain risk neutrality (aside for when choosing from a continuum of incentive compatible contracts) but also require an incentive to smooth consumption over time to create a meaningful intertemporal model. Therefore, we use the framework of Epstein-Zin (1990) and Weil (1989), which allows us to disentangle risk aversion and intertemporal

substitution.²⁹

Let W_t denote the CEO's wealth, and 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, $\ln V_t = \sum_{s=0}^{\infty} \delta^s ((1 - \delta) \ln c_{t+s} - \Lambda e_{t+s})$. Note that this still in essence a multiplicative model, like (1) and (23). The model is most suited for continuous time analysis, but for expositional clarity, we proceed in discrete time and take the continuous time limit where applicable.

With a logarithmic utility function, a standard result is that the indirect utility of wealth is $\ln V_t = \ln W_t + k$, where k is a constant independent of wealth. Therefore, the optimal consumption policy is $c_t = (1 - \delta) W_t$.

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

$$W_{t+1} = W_t (1 + r_f + \theta_t (e_t - \bar{e}) + \theta_t \eta_{t+1}) - c_{t+1}. \quad (33)$$

If the CEO shirks at time t , he increases his utility $\ln V_t$ by $\Lambda \bar{e} \Delta t$. On the other hand, his wealth at $t + 1$ is reduced by $\Delta W_t = -W_r(t) \bar{e} \Delta t$, where $W_r = \partial W / \partial r$. (In our example, $W_r = W \theta$.) Therefore, shirking increases utility $\ln V_t$ by:

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

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

$$\frac{\partial W}{\partial r} \geq \Lambda W \quad (34)$$

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

$$\frac{\partial W}{\partial r} = \Lambda W. \quad (35)$$

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

²⁹Risk neutrality significantly enhances tractability. Without smooth consumption, the model would be degenerate as the CEO consumes everything in a period in which he shirks.

C Options and Nonlinear Incentive Contracts

We return to the basic model with a binary effort decision, and generalize from stocks to a broader range of compensation instruments. The CEO receives fixed pay f , and ν units of a “security”; one unit of the security pays $V(P_1)$. For instance, for an option with strike K , $V(P_1) = \max(0, P_1 - K)$. Total compensation is $c = f + \nu V(P_1)$.

In equilibrium, the CEO should be paid $w \equiv E[cg(\bar{e}) | e = \bar{e}]$. If the CEO shirks ($e = \underline{e} \equiv 0$), the CEO’s utility is, calling P_1 the value of P_1 achieved if CEO exerts high effort.

$$\begin{aligned} E[cg(0) | e = 0] &= E[f + \nu V(P_1(1 - \bar{e}))]g(0) = E[f + \nu V(P_1) - \nu\Delta\bar{e}]g(0) \\ &= (w - \nu\Delta\bar{e}) / (1 - \Lambda\bar{e}), \end{aligned}$$

with

$$\Delta \equiv (E[V(P_1)] - E[V(P_1(1 - \bar{e}))]) / \bar{e}. \quad (36)$$

Hence, the CEO works if $E[cg(\bar{e}) | e = \bar{e}] \geq E[cg(0) | e = 0]$, i.e.

$$w \geq (w - \nu\Delta\bar{e}) / (1 - \Lambda\bar{e}) \Leftrightarrow \nu \geq \nu^* = w \frac{\Lambda}{\Delta}.$$

This leads to the following generalization of Proposition 1.

Proposition 10 (*General incentive contracts*). *Using nonlinear incentive contracts, the conclusions of Proposition 1 remain the same, with a change of notation. The manager’s expected pay is w , which comprises fixed base salary f^* , and $\nu^*E[V(P_1)]$ worth of securities, with:*

$$\begin{aligned} \text{Incentivized pay} &= \nu^*E[V(P_1)] = w\Lambda', \\ \text{Fixed pay} &= f^* = w(1 - \Lambda'), \end{aligned} \quad (37)$$

where $\Lambda' = \Lambda E[V(P_1)] / \Delta$, $\nu^* = w \frac{\Lambda}{\Delta}$, and Δ is in (36). Realized pay is:

$$c = w + \nu^*(V(P_1) - E[V(P_1)]).$$

Regressing the ex post compensation c on the firm return $r = P_1/P_0 - 1$ yields,

$$b^{III} = \frac{\partial E[c]}{\partial r} = \nu^*E\left[\frac{\partial V(P_0(1+r))}{\partial r}\right] = \nu^*P_0E[V'(P_1)] = w\frac{\Lambda}{\Delta}P_0E[V'(P_1)] = w\Lambda\xi,$$

with

$$\xi = \frac{P_0E[V'(P_1)]}{\Delta}. \quad (38)$$

For instance, if the security is a stock, $V(P) = P$, $\Delta = 1$, $E[V'(P_1)] = 1$, and $\xi = 1$. For small $P_1/P_0 - 1$ and \bar{e} , by Taylor expansion, $\Delta \rightarrow E[V'(P_1)]P_0$, and $\xi \rightarrow 1$. We can therefore

think of ξ as approaching 1, and so the broader economics are unchanged.

Proposition 11 *Using general incentive contracts, the conclusions of Proposition 3 remain the same, modified only by the introduction of a parameter ξ . The pay-performance sensitivities are:*

$$\begin{aligned} b_n^I &= \xi \frac{\Lambda}{L} \\ b_n^{II} &= \xi \frac{\Lambda w_n}{L S_n} \\ b_n^{III} &= \xi \frac{\Lambda}{L} w_n, \end{aligned}$$

with ξ given in (38). In many cases, $\xi \simeq 1$. Proposition 4 remains exactly the same.

D Detailed Calculation of B^I

We merge Compustat with ExecuComp (1992-2006) 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). If non-missing, we also subtract 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_volatility`. If it is missing, we replace it with the median volatility for that year. We also winsorize σ at the 5th and 95th percentile for each year.

r is the continuously compounded risk-free rate, available from <http://wrds1.wharton.upenn.edu/ds/comp/execcomp/means.html>.

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` (if this is missing, we set it equal to the stock price at the end of the fiscal year), and T can be calculated using the option’s maturity date, `exdate`. If `exdate` is unavailable, we assume a maturity of 9.5 years. (Standard options have a 10 year maturity; we assume the average option is granted mid-way through the year). 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 (`option_awards_num`) for the number of newly granted options, it is sometimes different from the number obtained by summing across grants, because `numsecur` is sometimes missing. 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

$$\text{prccf} - \frac{\text{opt_unex_unexer_est_val} - \text{ivnew}}{\text{opt_unex_unexer_num} - \text{numnewop}}.$$

`opt_unex_unexer_est_val` 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 $\max(0, (\text{prccf} - \text{expric})) * \text{numsecur}$ for each new grant and summing across new grants.

`opt_unex_unexer_num` 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

$$\text{prccf} - \frac{\text{opt_unex_exer_est_val}}{\text{opt_unex_exer_num}}.$$

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

`opt_unex_exer_num` is the number of exercisable options held at the end of the year.

In some cases, `numnewop` > `opt_unex_unexer_num`, i.e. the number of newly granted options exceeds the number of unexercisable options at year end. As in Core and Guay (2002),

we interpret these cases as part of the new grant ($\text{numnewop} - \text{opt_unex_unexer_num}$) being exercisable. We therefore calculate the strike price of exercisable options as

$$\text{prccf} - \frac{\text{opt_unex_exer_est_val}}{\text{opt_unex_exer_num} - (\text{numnewop} - \text{opt_unex_unexer_num})}.$$

In a subset of these cases, $\text{numnewop} > \text{opt_unex_unexer_num} + \text{opt_unex_exer_num}$, i.e. the number of newly granted options exceeds the number of total options at year end.³⁰ In such cases, we assume that some of the new options were exercisable, and that the CEO had already exercised them during the year.³¹

In some cases, $\text{ivnew} > \text{opt_unex_unexer_est_val}$, i.e. the intrinsic value of the newly granted options exceeds the intrinsic value of unexercisable options. In a subset of these cases, $\text{opt_unex_unexer_num} > \text{numnewop}$, i.e. there are some previously granted unexercisable options. We assume that such options are at the money. If $\text{ivnew} > \text{opt_unex_unexer_est_val}$ and $\text{numnewop} > \text{opt_unex_unexer_num}$, 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

$$\text{prccf} - \frac{\text{opt_unex_exer_est_val} - (\text{ivnew} - \text{opt_unex_unexer_est_val})}{\text{opt_unex_exer_num} - (\text{numnewop} - \text{opt_unex_unexer_num})}.$$

If $\text{ivnew} > \text{opt_unex_exer_est_val} + \text{opt_unex_unexer_est_val}$ but $\text{opt_unex_exer_num} > \text{numnewop} - \text{opt_unex_unexer_num}$ (i.e. there are some previously granted exercisable options), 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 new grants, we use 8.5 years.³² 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,

³⁰We checked selected cases against the original SEC form 14a filings. In some cases, this was due to inaccurate data entry by ExecuComp (in particular, ExecuComp reporting dollar rather than number amounts for the quantity of options). However, in other cases, ExecuComp reported accurately, hence the interpretation in the next sentence.

³¹Hence, we assume there are $\text{opt_unex_unexer_num}$ unexercisable options with a strike price of $\text{prccf} - \frac{\text{opt_unex_unexer_est_val}}{\text{opt_unex_unexer_num}}$, and opt_unex_exer_num exercisable options with a strike price of $\text{prccf} - \frac{\text{opt_unex_exer_est_val}}{\text{opt_unex_exer_num}}$.

³²Core and Guay (2002) recommend 9 years. We use 8.5 because we assume the the average new grant is given half-way through the fiscal year and thus has a maturity of 9.5 years.

exercisable options. In the very few cases where delta cannot be calculated because prccf is missing, we set deltaun and deltaex to 0.7 as estimated by Guay (1999).

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

$$\begin{aligned} \text{totaldelta} &= [\text{shrown} + \text{totaldeltanew} + \max(0, \text{opt_unex_unexer_num} - \text{numnewop}) \times \text{deltaun} \\ &\quad + \max(0, (\text{opt_unex_exer_num} - \max(0, \text{numnewop} - \text{opt_unex_unexer_num})) \\ &\quad \times \text{deltaex}] / 1000 \\ &= \frac{1}{P} \left[\text{Value of stock} + \text{Number of options} \times \frac{\partial V}{\partial P} \times P \right] \end{aligned}$$

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

We then calculate our measures of wealth-performance sensitivity, deflating all nominal variables using the GDP deflator from the Bureau of Economic Analysis website (<http://www.bea.gov/national/xls/gdplev.xls>):

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

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 omit 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:

$$\frac{W_{t+1} - W_t}{w_t} = A + \widehat{B}^I \times r_{t+1} + C \times r_{M,t+1} + \text{Controls}, \quad (39)$$

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).³³

Even if full wealth data (which includes flow compensation) is available, the ex ante measure

³³ $r_{M,t+1}$ 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 (39) 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.

has a number of advantages. First, both data on overall wealth and a long time series are required to estimate equation (39) 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.

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, University of Pennsylvania
- [3] Baker, G. and B. Hall, 2004, "CEO Incentives and Firm Size." *Journal of Labor Economics* 22, 767-98
- [4] Baranchuk, N., G. MacDonald and J. Yang, 2007, "The Economics of Super Managers." Working Paper, University of Texas at Dallas
- [5] Bebchuk, L. and J. Fried, 2004, *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press
- [6] Bebchuk, L., and Y. Grinstein, 2007, "Firm Expansion and CEO Pay," Working Paper, Harvard University
- [7] Becker, B., 2006, "Wealth and Executive Compensation." *Journal of Finance* 61, 379-397
- [8] Bennedsen, M. F. Perez-Gonzalez, and D. Wolfenzon, 2007, "Do CEOs Matter?" Working Paper, Copenhagen Business School
- [9] Carleton, W., J. Nelson and M. Weisbach, 1998, "The Influence of Institutions on Corporate Governance through Private Negotiations: Evidence from TIAA-CREF." *Journal of Finance* 53, 1335-1362
- [10] Coles, J.L., M. L. Lemmon and J. F. Meschke, 2007, "Structural Models and Endogeneity in Corporate Finance: The Link Between Managerial Ownership and Corporate Performance." Working Paper, Arizona State University
- [11] Cooley, T. and E. Prescott, 2005, "Economic Growth and Business Cycles," in "Frontiers in Business Cycle Research," T. Cooley ed., Princeton University Press, Princeton
- [12] 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
- [13] 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
- [14] Demsetz, H. and K. Lehn, 1985, "The Structure of Corporate Ownership: Causes and Consequences." *Journal of Political Economy* 93, 1155-1177

- [15] Dicks, D., 2007, "Executive Compensation, Incentives, and the Role for Corporate Governance Regulation." Working Paper, Northwestern University
- [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] Dittmann, I., E. Maug and O. Spalt, 2008, "Optimal CEO Compensation When Managers are Loss Averse." Working Paper, Erasmus University
- [18] Edmans, A., 2007, "Inside Debt." Working Paper, University of Pennsylvania
- [19] Edmans, A. and X. Gabaix, 2008, "Robust Incentive Contracting in Continuous Time." Working Paper, University of Pennsylvania
- [20] Edmans, A., I. Goldstein and W. Jiang, 2008, "Takeover Activity and Target Valuations: Evidence of Feedback in Financial Markets." Working Paper, University of Pennsylvania
- [21] 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
- [22] Falato, A., and D. Kadyrzhanova, 2007, "Optimal CEO Incentives and Industry Dynamics." Working Paper, Federal Reserve Board
- [23] Frydman, C. and R. Saks, 2007, "Historical Trends in Executive Compensation, 1936-2003." Working Paper, Massachusetts Institute of Technology
- [24] Gabaix, X. and A. Landier, 2008, "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123, 49-100
- [25] Garicano, L. and T. Hubbard, 2007, "The Return to Knowledge Hierarchies." NBER Working Paper No. 12815
- [26] Gayle, G. and R. Miller, 2007, "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" Working Paper, Carnegie-Mellon University
- [27] 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
- [28] Gompers, P., J. Ishii and A. Metrick, 2003, "Corporate Governance and Equity Prices." *Quarterly Journal of Economics* 118, 107-155
- [29] Grossman, S. and O. Hart, 1983, "An Analysis of the Principal-Agent Problem." *Econometrica* 51, 7-45

- [30] Guay, W., 1999, “The Sensitivity of CEO Wealth to Equity Risk: An Analysis of the Magnitude and Determinants.” *Journal of Financial Economics* 53, 43-78.
- [31] Habib, M. and A. Ljungqvist, 2005, “Firm Value and Managerial Incentives: A Stochastic Frontier Approach.” *Journal of Business* 78, 2053-2094
- [32] Hall, B. and J. Liebman, 1998, “Are CEOs Really Paid Like Bureaucrats?” *Quarterly Journal of Economics* 113, 653-691
- [33] Haubrich, J. G., 1994, “Risk Aversion, Performance Pay, and the Principal-Agent Problem.” *Journal of Political Economy* 102, 258-276.
- [34] He, Z., 2007, “Agency Problems, Firm Valuation, and Capital Structure,” Working Paper, Northwestern University
- [35] Hermalin, B. and M. Weisbach, 1998, “Endogenously Chosen Boards of Directors and Their Monitoring of Management.” *American Economic Review* 88, 96-118
- [36] Hermalin, B. and N. Wallace, 2001, “Firm Performance and Executive Compensation in the Savings and Loan Industry,” *Journal of Financial Economics* 61, 139–170
- [37] Holmstrom, B., 1992, “Comments,” in Lars Werin and Hans Wijkander, eds., *Contract Economics*. Blackwell, Cambridge MA
- [38] Holmstrom, B. and P. Milgrom, 1987, “Aggregation and Linearity in the Provision of Intertemporal Incentives.” *Econometrica* 55, 308-328
- [39] Inderst, R. and H. Mueller, 2007, “CEO Compensation and Private Information: An Optimal Contracting Perspective.” Working Paper, New York University
- [40] Jensen, M. and W. Meckling, 1976, “Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure.” *Journal of Financial Economics* 3, 305-360
- [41] Jensen, M. and K. J. Murphy, 1990, “Performance Pay and Top-Management Incentives.” *Journal of Political Economy* 98, 225-64
- [42] Kuhnen, C. and J. Zwiebel, 2007, “Executive Pay, Hidden Compensation and Managerial Entrenchment.” Working Paper, Northwestern University
- [43] Luttmer, E. G. J., “Selection, Growth, and the Size Distribution of Firms.” *Quarterly Journal of Economics*, 122, 1103-1144.
- [44] Manso, G., 2006, “Motivating Innovation.” Working Paper, Massachusetts Institute of Technology

- [45] Morck, R., A. Shleifer and R. Vishny, 1990, “Do Managerial Objectives Drive Bad Acquisitions?” *Rand Journal of Economics* 24, 357-379
- [46] Murphy, K. J., 1985, “Corporate Performance and Managerial Remuneration: An Empirical Analysis.” *Journal of Accounting and Economics*, 7, 11–42
- [47] 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
- [48] Oyer, P., 2004, “Why Do Firms Use Incentives That Have No Incentive Effects?” *Journal of Finance*, 59, 1619-1650
- [49] Prendergast, C., 2002, “The Tenuous Trade-Off between Risk and Incentives.” *Journal of Political Economy*, 110, 1071-102
- [50] Rajan, R. and J. Wulf, 2006, “Are Perks Purely Managerial Excess?” *Journal of Financial Economics* 79, 1-33
- [51] 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
- [52] Sattinger, M., 1993, “Assignment Models of the Distribution of Earnings,” *Journal of Economic Literature* 31, 831–880
- [53] Schaefer, S., 1998, “The Dependence of Pay-Performance Sensitivity on the Size of the Firm.” *Review of Economics and Statistics*, 80, 436-443
- [54] Taylor, L., 2007, “Why Are CEOs Rarely Fired? Evidence From Structural Estimation.” Working Paper, University of Chicago
- [55] Tervio, M., 2007, “The Difference that CEOs Make: An Assignment Model Approach.” Working Paper, University of California, Berkeley.
- [56] Weil, P., 1990, “Nonexpected Utility in Macroeconomics.” *Quarterly Journal of Economics* 105, 29-42
- [57] Yermack, D., 2006, “Flights of Fancy: Corporate Jets, CEO Perquisites, and Inferior Shareholder Returns”. *Journal of Financial Economics* 80, 211-242
- [58] Yermack, D., 1995, “Do Corporations Award CEO Stock Options Effectively?” *Journal of Financial Economics* 39, 237-269

Table 1: Comparing Different Measures of Incentives.

	b^I	b^{II}	b^{III}
PPS	$\frac{\Delta \ln c}{\Delta \ln S}$	$\frac{\Delta c}{\Delta S}$	$\frac{\Delta c}{\Delta \ln S}$
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) Gibbons-Murphy (1992) Rosen (1992)	Demsetz-Lehn (1985) Jensen-Murphy (1990) Yermack (1995) Schaefer (1998)	Holmstrom (1992) Hall-Liebman (1998)
This paper	Λ	$\Lambda \frac{w}{S}$	Λw
Scaling with S	$b^I \propto S^0$ $b^I \propto S^0$	$b^{II} \propto S^{\rho-1}$ $b^{II} \propto S^{-2/3}$	$b^{III} \propto S^\rho$ $b^{III} \propto S^{1/3}$
Scaling with $S(n_*)$	$b^I \propto S^0 S(n_*)^0$ $b^I \propto S^0 S(n_*)^0$	$b^{II} \propto S^{-(1-\rho)} S(n_*)^{\gamma-\rho}$ $b^{II} \propto S^{-2/3} S(n_*)^{2/3}$	$b^{III} \propto S^\rho S(n_*)^{\gamma-\rho}$ $b^{III} \propto S^{1/3} S(n_*)^{2/3}$

Explanation: This Table shows the three different measures of pay-performance sensitivity (PPS) and wealth-performance sensitivity (WPS). c is the realized compensation, w is the expected compensation, S is the aggregate value of the firm, W is CEO wealth, and 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 approximately $\rho = 1/3$. 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 Wealth-Performance Sensitivity with Firm Size.

	$\ln(B^I)$	$\ln(B^{II})$	$\ln(B^{III})$
$\ln(\text{Aggregate Value})$	0.0380 (0.0677)	-0.6023 (0.0520)	0.3977 (0.0520)
Year Fixed Effects	Yes	Yes	Yes
Industry Fixed Effects	Yes	Yes	Yes
Observations	6,470	6,470	6,470
Adj. R-squared	0.1323	0.3176	0.3444

Explanation: We merge Compustat with ExecuComp (1992-2006) 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 (24)-(26). The industries are the Fama-French (1997) 48 sectors. Standard errors, displayed in parentheses, are clustered at the firm level. The model predicts a coefficient 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 Wealth Volatility and Firm Volatility.

	Ex ante measure of volatility			Ex post measure of volatility		
	$\ln(B^I \sigma_r)$	$\ln(B^{II} \sigma_r)$	$\ln(B^{III} \sigma_r)$	$\ln\left(\frac{ W_{t+1}-W_t }{w_t}\right)$	$\ln\left(\frac{ W_{t+1}-W_t }{S_t}\right)$	$\ln W_{t+1} - W_t $
ln(Return Vol)	0.9979 (0.1305)	1.3632 (0.1196)	1.3632 (0.1196)	0.6831 (0.1633)	0.9885 (0.1389)	1.0635 (0.1423)
ln(Agg. Value)	0.0242 (0.0581)	-0.5848 (0.0479)	0.4152 (0.0479)	-0.0487 (0.0639)	-0.6202 (0.0512)	0.2960 (0.0523)
Year FE	Yes	Yes	Yes	Yes	Yes	Yes
Industry FE	Yes	Yes	Yes	Yes	Yes	Yes
Observations	6,276	6,276	6,276	4,545	4,545	4,545
Adj. R-squared	0.2238	0.4508	0.4508	0.1390	0.3200	0.2870

Explanation: We merge Compustat with ExecuComp (1992-2006) 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 (24)-(26). The industries are the Fama-French (1997) 48 sectors. Standard errors, displayed in parentheses, are clustered at the firm level. The model predicts a coefficient of 1 on $\ln(\text{Return Volatility})$, whereas models with unbounded effort predict a negative coefficient. The theory also predicts a coefficient on $\ln(\text{Aggregate Value})$ of $\rho = 0$ for B^I , $\rho = -2/3$ for B^{II} and $\rho = 1/3$ for B^{III} .