NBER WORKING PAPER SERIES

EXECUTIVE COMPENSATION:
A MODERN PRIMER

Alex Edmans
Xavier Gabaix

Working Paper 21131
http://www.nber.org/papers/w21131

NATIONAL BUREAU OF ECONOMIC RESEARCH
1050 Massachusetts Avenue
Cambridge, MA 02138
April 2015

We thank Steven Durlauf (the editor), two anonymous referees, Taylor Begley, David Dicks, Steve Kaplan, Robert Miller, Yuliy Sannikov, Fenghua Song, Luke Taylor, David Yermack, and especially Pierre Chaigneau for helpful comments, and Deepal Basak, Olivia Domba, and Jan Starmans for valuable research assistance. The views expressed herein are those of the authors and do not necessarily reflect the views of the National Bureau of Economic Research.

Executive Compensation: A Modern Primer
Alex Edmans and Xavier Gabaix
NBER Working Paper No. 21131
April 2015
JEL No. D86,G34

ABSTRACT

This article studies traditional and modern theories of executive compensation, bringing them together under a unifying framework. We analyze assignment models of the level of pay, and static and dynamic moral hazard models of incentives, and compare their predictions to empirical findings. We make two broad points. First, traditional optimal contracting theories find it difficult to explain the data, suggesting that compensation results from "rent extraction" by CEOs. In contrast, more modern theories that arguably better capture the CEO setting do deliver predictions consistent with observed practices, suggesting that these practices need not be inefficient. Second, seemingly innocuous features of the modeling setup, often made for tractability or convenience, can lead to significant differences in the model's implications and conclusions on the efficiency of observed practices. We close by highlighting apparent inefficiencies in executive compensation and additional directions for future research.

Alex Edmans
The Wharton School
University of Pennsylvania
2460 Steinberg Hall - Dietrich Hall
3620 Locust Walk
Philadelphia, PA 19104
and NBER
aedmans@wharton.upenn.edu

Xavier Gabaix
New York University
Finance Department
Stern School of Business
44 West 4th Street, 9th floor
New York, NY 10012
and NBER
xgabaix@stern.nyu.edu

1. Introduction

There is considerable debate on executive compensation in both the public arena and academia. This debate spans several important topics in economics, such as contract theory, corporate finance, corporate governance, labor economics, and income inequality. One side is the "rent extraction" view, which claims that current compensation practices sharply contrast the predictions of traditional optimal contracting models (e.g. Holmstrom (1979)). Thus, contracts are not chosen by boards to maximize shareholder value, but instead by the executives themselves to maximize their own rents. This perspective, espoused most prominently by Bebchuk and Fried (2004), has been taken very seriously by both scholars and policymakers, and led to major regulatory changes. In the U.S., the Securities and Exchange Commission ("SEC") mandated increased disclosure of compensation in 2006, and say-on-pay legislation was passed as part of Dodd-Frank in 2007. In 2013, the European Union capped bankers' bonuses at the level of their salary, or twice their salary if shareholders agree. In the same year, the SEC passed a rule requiring firms to disclose the ratio of Chief Executive Officer ("CEO") pay to median employee pay, and Switzerland held an ultimately unsuccessful referendum to limit CEO pay to twelve times the pay of the lowest worker.

The "optimal contracting" view reaches a different conclusion. While it acknowledges standard agency models are inconsistent with practice, it argues that such models do not capture the specifics of the CEO setting, since they were created as frameworks for the principal-agent problem in general. For example, CEOs have a very large effect on firm value compared to rank-and-file employees. Thus, in a competitive labor market, it may be optimal to pay high wages to attract talented CEOs, and implement high effort from them even though doing so requires paying a premium.[1] Newer optimal contracting models aim to capture the specifics of the CEO employment relationship, and do indeed generate predictions consistent with the data. Under this perspective, regulation will do more harm than good.

This article critically assesses the rent extraction vs. optimal contracting debate, by analyzing newer models of executive compensation and evaluating the extent to which they can explain observed practices. In particular, while recent theories have used different frameworks and focused on different dimensions of the contracting problem, we present a tractable unifying model to bring together the conclusions of this large literature, starting with classic theories and then moving to modern ones. In Section 2 we begin by analyzing the determinants of the level of pay, starting with neoclassical production models of the firm and then moving to modern assignment models. These assignment models yields empirical predictions for how CEO pay varies cross-sectionally between firms of different sizes, and over time as the size of the average firm in the economy changes.

Having determined the level of pay, we then move to incentives. In Section 3 we consider a

[1] A simple model can justify high CEO pay simply by assuming a high level for the reservation utility, which is an exogenous parameter. Modern assignment models endogenize the reservation utility.

static moral hazard model where the CEO takes an action that improves expected firm value, starting in Section 3.1 with the risk-neutral case and moving to risk aversion in Section 3.2. While this setting appears quite standard, we will show that seemingly innocuous features of the modeling setup, often made for convenience or tractability (e.g. the choice between additive or multiplicative utility and production functions, and binary or continuous actions) can lead to significant differences in the model's implications – and thus conclusions as to whether observed practices are consistent with theory. We also discuss various frameworks that researchers can use to yield tractable solutions to the contracting problem, and the appropriate empirical measure of incentives. Section 3.3 embeds the moral hazard model into a market equilibrium to generate additional empirical implications, and Section 3.4 discusses the evidence. Section 3.5 considers the case of multiple signals. In contrast to the Holmstrom (1979) informativeness principle, CEO pay in reality depends on industry shocks outside his control, which Bebchuk and Fried (2004) argue is strong evidence that contracting is suboptimal. We show that the theory does not unambiguously predict that industry shocks should be filtered out due to other considerations in a CEO setting that are absent from Holmstrom (1979). Section 3.6 allows the CEO to affect the volatility as well as mean of firm value, by choosing the firm's risk. It discusses how options can encourage "good" risk-taking, and debt-based compensation can deter "bad" risk-shifting if the firm is levered.

Section 4 moves to a multi-period model. A dynamic setting poses several challenges absent from a single-period model: contracts that are initially optimal may lose their incentive effect over time, the CEO can take myopic actions that boost short-term returns at the expense of long-run value, and he may undo the contract by private saving. In addition to these complications, a dynamic setting provides the principal with additional opportunities: she can provide incentives through the threat of termination, and base the CEO's pay on returns in previous as well as current periods.

Each section will compare the empirical predictions of the theories with the evidence. Broadly speaking, we will argue that many, but not all, features of observed contracts that are frequently criticized are actually consistent with efficiency. However, empirical correlations cannot be interpreted as definitive proof of the optimal contracting view, given the difficulties in identifying causality. Section 5 highlights apparent inefficiencies in executive compensation, as well as open questions for future research. Section 6 concludes.

This article aims to differ from existing surveys of executive compensation. Core, Guay, and Larcker (2003), Jensen and Murphy (2004) and Frydman and Jenter (2010) focus largely on the empirical evidence. Murphy (2013) provides a historical perspective and discusses the role of institutional constraints. Edmans and Gabaix (2009) focus exclusively on recent theories and use verbal descriptions rather than a formal model. Our main contribution is to study both traditional and modern contracting theories, with a particular attention to the role of modeling choices, and combine their findings into a single unifying framework. As with any survey, we are forced to draw boundaries and thus the analysis of asymmetric information focuses on moral hazard rather than adverse selection as the former literature is more extensive. For learning models of CEO contracts, we refer the interested reader to Harris and Holmstrom (1982), Gibbons and Murphy (1992), Holmstrom (1999), Hermalin and Weisbach (1998, 2012), Taylor (2010, 2013), and Garrett and Pavan (2012).

2. The Level of Pay

Trends in the level of pay are perhaps the most commonly cited statistic in support of the rent extraction view. The median CEO in the S&P 500 earned $9.6 million in 2011 (Murphy (2013)), which is substantially higher than in other countries and represents a sixfold increase since 1980. In contrast, the pay of the average worker has risen much more slowly. Figures from the Bureau of Labor Statistics show that CEO pay was 350 times that of the average worker in 2013, compared to 40 times in 1980 according to the Economic Policy Institute. Thus, the rapid increase in executive compensation may have contributed significantly to the recent rise in income inequality (Piketty and Saez (2003), Piketty (2014)), and has potential political economy implications. Bebchuk and Fried (2004) argue that this increase is a result of rent extraction by CEOs. Supporting this argument, Bebchuk, Cremers, and Peyer (2011) show that the fraction of CEO pay relative to total pay across the top-five executives is linked to lower firm value, profitability, and returns to acquisition announcements. We study the extent to which rises in pay over time can be explained by optimal contracting models. In this section, we abstract from agency problems (which later introduce in Section 3) and study the pay required to attract the CEO to a firm.

2.1. Talent as a factor of production

One approach to determining the level of pay is to view the CEO as a factor of production separate from standard employees. Let the production function be

$$V = F(K, L, T),$$

where V is firm value and the factors of production are units of capital K, number of workers L ("labor"), and number of managers T. Each manager is paid a wage $w_T = \frac{\partial F}{\partial T}$: his pay is determined by the production function, and changes in pay result from shifts in technology. This is the perspective of most economic theories on the aggregate production function and supply of talent (see Goldin and Katz (2009) and Acemoglu and Autor (2012) for recent surveys). In particular, labor economists use this perspective to compare the wages of, say, college graduates vs. high-school dropouts.

The Lucas (1978) theory of the firm specializes the model to apply to the pay of a single CEO, rather than several managers. The variable T now refers to the CEO's level of human capital (i.e. his talent) rather than the number of managers. A CEO with talent T hires capital and labor[2], and maximizes:

$$W_T = \max_{K,L} F(K, L, T) - w_L L - rK, \tag{1}$$

where w_L and r are the prices of labor and capital, and the surplus W_T is the CEO's pay. Consider the Cobb-Douglas production function

$$V = T^{\alpha_T} \left(\frac{K}{\alpha_K}\right)^{\alpha_K} \left(\frac{L}{\alpha_L}\right)^{\alpha_L}, \tag{2}$$

where α_T, α_K, and α_L represent the shares of output that go to the CEO, capital, and labor, respectively, under perfect competition. We assume $\alpha_T + \alpha_K + \alpha_L = 1$ (constant returns to scale). The first-order condition of (1) with respect to K is $\alpha_K \frac{V}{K} = r$ i.e. $\frac{K}{\alpha_K} = \frac{V}{r}$. Optimizing over labor likewise yields $\frac{L}{\alpha_L} = \frac{V}{w}$. Substituting into the production function (2) gives:

$$V = T^{\alpha_T} \left(\frac{V}{r}\right)^{\alpha_K} \left(\frac{V}{w_L}\right)^{\alpha_L} = \frac{1}{r^{\alpha_K} w_L^{\alpha_L}} T^{\alpha_T} V^{1-\alpha_T}.$$

Solving for V, we have:

$$\begin{aligned} V &= \left(r^{\alpha_K} w^{\alpha_L}\right)^{-1/\alpha_T} T \\ K &= \frac{\alpha_K V}{r}, L = \frac{\alpha_L V}{w_L}. \end{aligned} \tag{3}$$

From (3), a more talented CEO runs a larger firm, in part because he hires more capital and labor: V, K, L are all linear in T. His pay is given by

$$W_T = V - rK - w_L K = \alpha_T V. \tag{4}$$

The model generates the qualitative prediction that CEO pay, W_T, is increasing in firm size, because a larger firm generates more surplus. It also generates the quantitative prediction that his pay scales linearly with firm size. Various empirical studies confirm the qualitative prediction that CEO pay is increasing in firm size[3]: Baker, Jensen and Murphy (1988, p.609) call this relationship "the best documented empirical regularity regarding levels of executive compensation." However, the quantitative prediction that pay is linear in firm size is contradicted by the data. The above papers find that CEO pay increases as a power function of firm size $W_T \sim S^\kappa$, where a typical elasticity is $\kappa \simeq 1/3$. Hence, the Lucas model needs to be refined. This is what assignment models do, to which we now turn.

[2] An alternative formulation is for capital to hire the CEO and labor. In competitive markets, the identity of the principal is immaterial. Here, we follow the Lucas (1978) formulation for ease of exposition.

[3] See, e.g., Roberts (1956), Baker, Jensen and Murphy (1988), Barro and Barro (1990), Cosh (1975), Frydman and Saks (2005), Joskow et al. (1993), Kostiuk (1990), Rose and Shepard (1997), and Rosen (1992).

2.2. Assignment models

Gabaix and Landier (2008) present a market equilibrium model of CEO pay. A continuum of firms and potential CEOs are matched together. Firm $n \in [0, N]$ has a "baseline" size $S(n)$ and CEO $m \in [0, N]$ has talent $T(m)$. Low n denotes a larger firm and low m a more talented CEO: $S'(n) < 0$, $T'(m) < 0$. n (m) can be thought of as the rank of the firm (CEO), or a number proportional to it, such as its quantile of rank.

We consider the problem faced by one particular firm. At $t = 0$, it hires a CEO of talent $T(m)$ for one period. The CEO's talent increases firm value according to

$$V = S(n) + CT(m)S(n)^{\gamma}, \tag{5}$$

where C parametrizes the productivity of talent and γ the elasticity of talent with respect to firm size. If $\gamma = (<) 1$, the model exhibits constant (decreasing) returns to scale.

We now determine equilibrium wages, which requires us to allocate one CEO to each firm. Let $w(m)$ denote the equilibrium wage of a CEO with index m. Firm n, taking the market wage of CEOs as given, selects CEO 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), Terviö (2008)):

$$w(n) = -\int_{n}^{N} CS(u)^{\gamma} T'(u)\, du + \underline{w}_N. \tag{6}$$

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, Gabaix and Landier (2008) use the following asymptotic value for the spacings of the talent distribution: $T'(n) = -Bn^{\beta-1}$. These functional forms give the wage in closed form, taking the limit as $n/N \to 0$:

$$w(n) = \int_{n}^{N} A^{\gamma} BCu^{-\alpha\gamma+\beta-1} du + \underline{w}_N = \frac{A^{\gamma} BC}{\alpha\gamma - \beta}\left[n^{-(\alpha\gamma-\beta)} - N^{-(\alpha\gamma-\beta)}\right] + \underline{w}_N \sim \frac{A^{\gamma} BC}{\alpha\gamma - \beta} n^{-(\alpha\gamma-\beta)}. \tag{7}$$

To interpret equation (7), we consider a reference firm, for instance the median firm in the universe of the top 500 firms. Denote its index n_*, and its size $S(n_*) = An_*^{-\alpha}$. Using

$S(n) = An^{-\alpha}$, we derive:

$$w(n) = \frac{A^{\gamma}BC}{\alpha\gamma - \beta} n^{-(\alpha\gamma-\beta)} = \frac{A^{\gamma}BC}{\alpha\gamma - \beta} \left(\left(A^{1/\alpha} S(n)^{-1/\alpha} \right) \right)^{-(\alpha\gamma-\beta)}$$
$$= \frac{A^{\beta/\alpha}BC}{\alpha\gamma - \beta} S(n)^{\gamma-\beta/\alpha} = \frac{\left(S(n_*) n_*^{\alpha}\right)^{\beta/\alpha} BC}{\alpha\gamma - \beta} S(n)^{\gamma-\beta/\alpha} = \frac{n_*^{\beta}BC}{\alpha\gamma - \beta} S(n_*)^{\beta/\alpha} S(n)^{\gamma-\beta/\alpha}.$$

Finally, we obtain the "dual scaling equation" for CEO pay:

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

where $D(n_*) = -Cn_* T'(n_*) / (\alpha\gamma - \beta)$ is a constant independent of firm size. Similar to Lucas (1978), equation (8) yields the qualitative cross-sectional prediction that CEO pay is increasing in firm size. However, the intuition is different: here the prediction arises because large firms hire the most talented CEOs, who command the highest wages. Moreover, equation (8) yields a different quantitative prediction. It predicts a pay-firm size elasticity of $\rho = \gamma - \beta/\alpha$. Gabaix and Landier (2008) calibrate using $\alpha = 1$ (a Zipf's law for firms, as in Axtell (2001)) and $\gamma = 1$ (constant returns to scale). Since there is no clear a priori value for β, they set $\beta = 2/3$ to yield the empirical pay-size elasticity of $\rho = 1/3$, which contrasts Lucas's (1978) prediction of $\rho = 1$. Baranchuk, MacDonald, and Yang (2011) extend the model to endogenize firm size and show that the pay-size relationship is stronger when industry conditions are favorable, as talented CEOs are not only paid a greater premium but also optimally grow their firms to a larger size.

In addition, equation (8) shows that pay increases with the size of the average firm in the economy $S(n_*)$. Since a CEO's talent can be applied to the entire firm, when firms are larger, the benefits from a more talented CEO are higher in dollar terms, and so there is more competition for talent. This is a similar "superstars" effect to Rosen (1992). Moreover, the model's closed form solutions yield quantitative predictions. Average firm size increased sixfold between 1980 and 2011. When both $S(n_*)$ and $S(n)$ rise by a factor of 6, CEO pay should rise by a factor of $6 \times [\beta/\alpha + (\gamma - \beta/\alpha)] = 6\gamma = 6$, as has been the case.[4] The relevant benchmark against which to compare the level of CEO pay is not the pay of the average worker, or pay of CEOs in the past, but his current contribution to the firm. This in turn depends on variables such as firm size and talent; while the latter is difficult to measure, the pay of the average worker is unlikely to be a determinant. Thus, assignment models suggest that regulation to mandate disclosure of the ratio of CEO pay to median employee pay may not be useful, as median employee pay is not the relevant benchmark.

However, the empirical evidence is not unambiguously in favor of assignment models. Nagel (2010) raises sample selection concerns and suggests alternative methodologies, but Gabaix, Landier, and Sauvagnat (2014) conclude that the results are robust to these changes. While

[4] Gabaix, Landier, and Sauvagnat (2014) provide an empirical update, with slight methodological changes. The main conclusions are robust to those changes.

Gabaix and Landier (2008) can fully explain the growth in CEO pay from 1980 to the present, Frydman and Saks (2010) find that median CEO pay was relatively constant between the 1940s and early 1970s despite firm size increasing over this period. Gabaix and Landier (2008) discuss potential explanations for this apparent discrepancy. One is that the supply of talent greatly increased, which creates a downward pressure on CEO wages; quantifying the impact of increased supply on wages would be a useful direction for future research. Another possibility is that the U.S. CEO market in that early period was similar to the Japan CEO market today, where CEOs are typically developed in-house rather than hired externally.

In addition, because both firm size and CEO pay have trended upwards since the 1970s, it is difficult to interpret the relationship as causal. Even if causal, the positive correlation between pay and firm size cannot be interpreted as definitive evidence in support of assignment models, since it is also potentially explainable by an (as yet unwritten) rent extraction model. For example, large firms may have more resources, allowing the CEO is able to extract more salary. Alternatively, Dicks (2012) shows that the correlation can arise if poor governance causes a small fraction of firms to overpay for talent, which then forces all others to overpay as well in order to remain competitive. This channel is also predicted by Gabaix and Landier (2008); see Bereskin and Cicero (2013) for supportive evidence. While these alternative explanations would generate the qualitative prediction that pay is positively correlated with firm size and average firm size, it is not yet clear whether they can generate empirically consistent quantitative predictions.

Another concern is that assignment models predict a reassignment of CEOs as relative firm size changes. While external poaching of CEOs has increased in recent years, it is still relatively rare: Cremers and Grinstein (2015) find that 63% of new CEOs are insiders. Similarly, it does not appear to be the case that large changes in firm size (e.g. a firm undertaking a large acquisition) are accompanied by changes in the acquiring CEO for non-disciplinary reasons. This low mobility can be generated by a simple extension of assignment models to incorporate frictions, such as a cost of firing the CEO or firm-specific human capital.[5] Writing and calibrating such a model would be valuable.

The level of CEO pay has implications not only for shareholders of the firm in question, but also the broader issue of income inequality: Piketty (2014, p. 302) finds that "the vast majority (60 to 70 percent, depending on what definitions one chooses) of the top 0.1 percent of the income hierarchy in 2000-2010 consists of top managers." Kaplan and Rauh (2010) put the rise in CEO pay in the context of income growth in other highly-paid occupations such as law, hedge funds, private equity, and venture capital. They find that CEO pay in public firms has risen significantly more slowly than these other occupations. Bakija, Cole, and Heim (2012) use restricted data from income tax statistics and confirm these results for public firms

[5]In addition, Cremers and Grinstein (2015) find the relationship between pay and firm size differs little across industries according to the proportion of outsider CEOs in the industry. They interpret this proportion as a potential measure of competition for CEO talent. However, this proportion may reflect small frictions that cause a firm to choose an insider CEO on the margin, but are not large enough to meaningfully affect equilibrium pay.

– for example, the percentage of non-finance salaried executives in the top 1% of the income distribution fell from 21.0% in 1979 to 11.3% in 2005. Across their overall sample, which mostly contains private firms, they find that executives, managers, and financial professionals account for 70% of the increase in the share of national income going to the top 0.1% between 1979 and 2005. Since agency problems are likely much lower in private firms due to close monitoring by the private equity owner, one interpretation is that agency problems may have not contributed to increased income inequality. However, such a conclusion is tentative without a model, since general equilibrium effects can be important for the overall level of pay.

Moving to other explanations for high CEO pay, Section 3 discusses how agency problems may lead to the CEO being paid a premium for the disutility of effort and the risk imposed by incentives. Other papers point to the changing nature of the employment relationship. Hermalin (2005) argues that tighter corporate governance increases both the level of effort that the CEO must exert and the risk of dismissal, and so managers demand greater pay as a compensating differential. Indeed, Peters and Wagner (2014) show that CEO turnover risk is significantly positively associated with pay, but reject an entrenchment model in which powerful CEOs enjoy both lower turnover risk and high pay. Their identification strategy focuses on industry volatility, which (after controls) is unlikely to affect CEO pay other than through turnover risk. In Garicano and Rossi-Hansberg (2006), CEOs specialize in knowledge acquisition and problem solving, leaving routine production tasks to lower-level employees. Recent increases in communication technologies (e.g. email) allow the CEO to specialize more on skilled tasks, thus increasing his pay. Frydman (2013) and Murphy and Zabojnik (2007) argue that the increasing importance of transferable, rather than firm-specific, human capital increases pay through expanding CEOs' outside options.

The above explanations – talent, agency, and the changing nature of the CEO's job – are part of the optimal contracting view. To test this view more broadly, Cronqvist and Fahlenbrach (2013) study the effect on compensation of firms transitioning from public to private ownership. Since the private equity sponsor's concentrated stake gives it the incentives and control rights to set pay optimally, compensation in private firms should be closer to the optimal contracting benchmark. Salary and bonuses actually rise upon going private, inconsistent with the notion that CEOs of public firms are overpaid. Indeed, Kaplan (2012) reports that over the past three decades, executive pay in closely-held firms has outpaced that in public companies.

In addition to the optimal contracting and rent extraction views, a third perspective is that institutional constraints or practices may have contributed to the rise of pay. Murphy (2013) discusses the role of tax policy, accounting rules, and disclosure requirements – for example, the Clinton Administration's $1 million salary cap led to many firms increasing salary to $1 million. Shue and Townsend (2015) note that firms tend to grant the same number of options each year. Thus, when stock prices rise, the value of options increases which, together with downward rigidity in salaries and bonuses, leads to overall pay levels rising in the 1990s and early 2000s. While this friction is indeed significant in the short run, its effect on long-run outcomes is less clear – similar to economics more broadly, where pricing frictions (e.g. menu costs) are important in the short-run but less so in the long-run.

3. Static Incentives

We now turn from determining the level of pay to the CEO's incentives. This section considers a single-period moral hazard model, which we extend to multiple periods in Section 4. This setting has been widely covered in textbooks (e.g. Bolton and Dewatripont (2005), Tirole (2005)) and earlier surveys (Prendergast (1999)), but typically with additive production functions and preferences, and often a binary effort level. We will show that multiplicative specifications, which may be particularly relevant for a CEO setting, lead to quite different conclusions for the best empirical measure of incentives and how incentives should vary cross-sectionally between firms. We will also show that the use of a continuous vs. binary action space, as well as the specification of noise before vs. after the action, also lead to significant differences in the model's results.

We start with a standard principal-agent problem applied to an executive compensation setting. The principal (board of directors on behalf of shareholders) hires an agent (CEO) to run the firm. The production function is given by $V(a, S, \varepsilon)$, which is increasing in a and S. We specialize this to

$$V = S + b(S)\,a + \varepsilon. \tag{9}$$

We consider an all-equity firm for simplicity and discuss leverage in Section 3.6. The variable $a \in [0, \infty)$ is an action taken by the agent that improves expected firm value but is personally costly. Examples include effort (low a represents shirking), project choice (low a involves selecting value-destructive projects that maximize private benefits), or rent extraction (low a reflects cash flow diversion.) We typically refer to a as "effort" for brevity. The variable ε is mean-zero noise, with interval support on $(\underline{\varepsilon}, \overline{\varepsilon})$, where the bounds may be infinite.[6] Shortly after the agent takes his action, noise is realized, and then final firm value V is realized. Firm value is observable and contractible, but neither effort nor noise are individually observable.

The function $b(S)$ measures the effect of effort on firm value for a firm of size S. One possibility is $b(S) = b$, which yields $V(a) = S + ba + \varepsilon$: an additive production function where the effect of effort on firm value is independent of initial firm size. This specification is appropriate for a perk consumption decision, if the amount of perks that can be consumed is independent of firm size. For example, buying a \$10 million corporate jet reduces firm value by \$10 million, regardless of S. Another is $b(S) = bS$, which yields $V(a) = S(1 + ba) + \varepsilon$: a multiplicative production function where the effect of firm value is linear in firm size. Many CEO actions can be "rolled out" across the entire firm and thus have a greater effect in a

[6] For simplicity, we assume that S is sufficiently large, or the probability of low ε is sufficiently small, that V is non-negative almost surely and so we do not need to complicate the model with non-negativity constraints.

larger company, such as a change in strategy or a program to improve production efficiency. A multiplicative specification is also appropriate for a rent extraction setting, if there are greater resources to extract in a larger firm.[7]

The agent is paid a wage $c(V)$ contingent upon firm value. (Note that c refers to actual pay, in contrast to w which refers to the expected wage). We always assume limited liability on the principal ($c(V) \leq V$): she cannot pay out more than total firm value. In some versions of the model we will also assume limited liability on the agent ($c(V) \geq 0$). He has reservation utility of $w \geq 0$ and his objective function is given by:

$$\mathrm{E}[U] = \mathrm{E}[u(v(c) - g(a))]. \tag{10}$$

The function g represents the cost of effort, which is increasing and weakly convex. u is the utility function and v is the felicity[8] function which denotes the agent's utility from cash; both are increasing and weakly concave. g, u, and v are all twice continuously differentiable. The objective function (10) contains functions for both utility and felicity to maximize generality. One common assumption is $v(c) = c$ so that $\mathrm{E}[U] = \mathrm{E}[u(c - g(a))]$, in which case the cost of effort is pecuniary, i.e. can be expressed as a subtraction to cash pay. This is appropriate if effort involves a financial expenditure or the opportunity cost of forgoing an alternative income-generating activity. Another is $u(x) = x$, which yields $\mathrm{E}[v(c) - g(a)]$, where the cost of effort is separable from the benefits of cash. This specification is reasonable if effort involves disutility, or forgoing leisure or private benefits.

Both of the above specifications represent additive preferences. Effort of a reduces the agent's utility by $\frac{1}{2}ga^2$ in utils (dollars) in the first (second) specification. A third specification is $v(c) = \ln c$, in which case (10) becomes $\mathrm{E}\left[u\left(\ln\left(ce^{-g(a)}\right)\right)\right]$. This specification corresponds to multiplicative preferences, where the cost of effort is increasing in c. Here, private benefits are a normal good: the utility they provide is increasing in consumption, consistent with the treatment of most goods and services in consumer theory. This specification is also plausible under the literal interpretation of effort as forgoing leisure: a day of vacation is more valuable to a richer CEO, as he has more wealth to enjoy during it. Thus, the CEO's expenditure on leisure and private benefits rises in proportion to his wealth. Multiplicative preferences are also commonly used in macroeconomic models (e.g. Cooley and Prescott (1995)) to generate realistic income effects. In particular, they are necessary for labor supply to be constant over time as the hourly wage rises.[9]

[7] See Bennedsen, Perez-Gonzalez, and Wolfenzon (2010) for empirical evidence that CEOs have the same percentage effect on firm value regardless of firm size.

[8] We note that the term "felicity" is typically used to denote one-period utility in an intertemporal model. We use it in a non-standard manner here to distinguish it from the utility function u.

[9] When the hourly wage rises, working becomes preferable to leisure (the substitution effect). With multiplicative preferences, the rise in the wage increases the agent's labor endowment income and thus demand for leisure (the income effect), which exactly offsets the substitution effect. With additive preferences, there is no income effect, and so leisure falls to zero as the wage increases.

The principal is assumed to be risk-neutral, since shareholders are typically well-diversified. Her program is given by:

$$\max_{c(\cdot),a} \mathrm{E}\left[V(a) - c(V(a))\right] \text{ s.t.} \tag{11}$$

$$\mathrm{E}\left[u\left(v\left(c\left(V(a)\right)\right) - g(a)\right)\right] \geq w \tag{12}$$

$$a \in \arg\max_{\widehat{a}} \mathrm{E}\left[u\left(v\left(c\left(V(\widehat{a})\right)\right) - g(\widehat{a})\right)\right]. \tag{13}$$

She chooses the effort level a and contract $c(V)$[10] to maximize (11), expected firm value minus the expected wage, subject to the agent's individual rationality or participation constraint ("IR", (12)) and incentive compatibility constraint ("IC", (13)).

We begin with a first-best benchmark, which leads to a simple optimal contract that is the same across all firms and thus does not have the potential to explain observed contracts. We then explore two departures from the first-best which generate a meaningful contract that does yield empirical predictions. The first is limited liability, which only leads to small variations in the optimal contract across firms. The second is risk aversion (Section 3.2) which leads to much richer implications. Section 3.3 embeds the contracting problem in a market equilibrium to generate additional empirical implications. We compare all implications to the data in Section 3.4.

Under the first-best, effort is observable. Let a^* be the effort level that the principal wishes to implement. She can simply direct the agent to exert effort a^*, and so we can ignore the IC (13). It is easy to show that the agent is given a constant wage $c(V) = \overline{c}$, as this leads to efficient risk-sharing. The IR (12) yields $\overline{c} \geq w + g(a^*)$. This will bind in the optimal contract, and so the principal maximizes

$$\mathrm{E}\left[V(a^*)\right] - g(a^*) - w. \tag{14}$$

This defines the first-best effort level as

$$g'\left(a^*_{FB}\right) = b(S). \tag{15}$$

The principal trades off the marginal increase in firm value from effort, $b(S)$, with the agent's marginal cost, $g'(a^*_{FB})$. Thus, a^*_{FB} maximizes total surplus. In turn, a^*_{FB} is decreasing in the convexity of the cost of effort. It is also increasing in firm size S if $b(S)$ is increasing in S, since effort then has a greater dollar effect in a larger firm.

We now turn to a setting in which effort is unobservable and the IC (13) must be imposed. We first assume a risk-neutral agent, before moving to risk aversion.

[10] Here, we focus on deterministic contracts, so that there is a one-to-one mapping between firm value V and compensation c. An even more general model allows for stochastic contracts, where firm value of V leads to a random amount c. Gjesdal (1982), Arnott and Stiglitz (1988), and Edmans and Gabaix (2011b) derive sufficient conditions for random contracts to be suboptimal, allowing the focus on deterministic contracts.

3.1. Risk-Neutral Agent

We first consider risk neutrality and additive preferences. We have $u(x) = x$ and $v(c) = c$ so the participation and incentive constraints (12) and (13) specialize to

$$\mathrm{E}[c(V)] - g(a) \geq w \tag{16}$$

$$a \in \arg\max_{\widehat{a}} \mathrm{E}[c(V)] - g(\widehat{a}). \tag{17}$$

Grossman and Hart (1983) show that the contracting problem can be solved in two stages, which correspond to the principal's two choice variables. She first chooses the contract $c(V)$ that implements a given action a^* at least cost, and then the optimal a^* taking into account the cost of the contract $c(V)$ needed to implement each action a^*. Starting with the first stage, the first-order condition of the agent's effort choice (17) is given by

$$\mathrm{E}[c'(V)\, b(S)] = g'(a^*). \tag{18}$$

Rogerson (1985), Jewitt (1988), and Carroll (2012) give conditions under which the first-order condition is sufficient, and so the IC (17) can be replaced by the first-order condition (18), which greatly simplifies the problem. Throughout this paper, we assume that these conditions are satisfied, so that the first-order approach is valid.

Given risk neutrality and unlimited liability, there is no loss of generality in focusing on a linear contract of the form $c(V) = \phi + \theta V$, where ϕ is the fixed wage and θ is the agent's percentage stake in firm value. Then, using (18), optimal incentives are given by

$$\theta = \frac{g'(a^*)}{b(S)}. \tag{19}$$

and so the implemented effort level is given by

$$a^* = g'^{-1}(\theta b(S)).$$

Empiricists typically measure the CEO's incentives to improve firm value, i.e. to exert effort a. Equation (19) shows how the optimal measure of incentives depends on how we specify the production function. When it is additive ($b(S) = b$), then we have $a^* = g'^{-1}(\theta b)$. The optimal measure of incentives is θ, the agent's percentage stake in firm value V. This measure corresponds to the dollar change in pay for a one dollar change in firm value ("\$-\$ incentives") and is used by Demsetz and Lehn (1985) and Jensen and Murphy (1990), among others.

Hall and Liebman (1998) argue that most CEO actions have a multiplicative effect on firm value. With a multiplicative production function ($b(S) = bS$), we have $a^* = g'^{-1}(\theta bS)$, in which case the optimal measure is θS, the CEO's dollar equity stake. This measure corresponds to the dollar change in pay for a one percentage point change in firm value ("\$-% incentives"). Thus, while it is common to assume an additive production function for simplicity, researchers should think carefully about how to specify these functions as this choice has important implications for the relevant measure of incentives – a point first noted by Baker and Hall (2004). Moreover, if CEO effort has a multiplicative effect on firm value, then CEO incentives are a quantitatively much more important issue than his level of pay, even though the latter receives much greater attention in the media. While a \$9.6 million salary is substantial compared to average worker pay, relative to a \$10 billion firm it constitutes 0.1% of firm value. In contrast, if incentives are insufficient to induce the CEO to implement a major restructuring or reject a bad acquisition, the losses to shareholders could run into several percentage points.

Before moving to the second stage of Grossman and Hart (1983), we demonstrate the effect of multiplicative preferences, as studied by Edmans, Gabaix, and Landier (2009), while retaining risk neutrality for now. In the general objective function (10), this corresponds to $u(x) = e^x$ and $v(c) = \ln c$, which yields

$$U = E\left[ce^{-g(a)}\right].$$

We normalize $a^* = 0$, and so the $t = 0$ stock price (net of CEO pay) is S.[11] In (9) we have $b(S) = bS$, i.e. a multiplicative production function, so that firm value at $t = 1$ is given by

$$V(a) = S(1 + ba) + \varepsilon.$$

The IR is given by $E[c|a = a^*] = w$, which yields:

$$w = [c|a = a^*] = \phi + \theta E[V|a = a^*] = \phi + \theta bS.$$

If the CEO exerts effort a, his utility is:

$$\begin{aligned} \mathrm{E}[U(a)] &= \mathrm{E}\left[c(a)\, e^{-g(a)}\right] = (\phi + \theta\, \mathrm{E}\, V(a))\, e^{-g(a)} \\ &= (\phi + \theta S(1 + ba))\, e^{-g(a)} = (w + \theta Sba)\, e^{-g(a)} \\ &= w\left(1 + \frac{\theta Sb}{w} a\right) e^{-g(a)} = we^{\ln\left(1 + \frac{\theta Sb}{w} a\right) - g(a)}. \end{aligned}$$

The IC is $a^* \in \arg\max_a \mathrm{E}[U(a)]$. At $a^* = 0$, this yields $EU'(0) = 0$, i.e.

$$\theta = \frac{g'(a^*)\, w}{bS}, \tag{20}$$

and so the implemented effort level is given by

$$a^* = g'^{-1}\left(\frac{\theta bS}{w}\right).$$

[11] For simplicity, we assume that initial firm size S is net of the expected wage w.

The model implies that the relevant measure of incentives is $\frac{\theta S}{w}$, i.e. the CEO's dollar equity stake scaled by his annual pay, or alternatively the fraction of total pay w that is in equity. Edmans, Gabaix, and Landier (2009) call this measure "scaled wealth-performance sensitivity". It corresponds to the percentage change in pay for a one percentage point change in firm value ("%-% incentives", i.e. the elasticity of pay to firm value), as used by Murphy (1985), Gibbons and Murphy (1992), and Rosen (1992).

Using θ^I, θ^{II}, and θ^{III}, respectively, to denote %-%, \$-\$, and \$-% incentives, we have:

$$\theta^I = \frac{\partial c}{\partial r}\frac{1}{w} = \frac{\Delta \ln \text{Pay}}{\Delta \ln \text{Firm Value}} \tag{21}$$

$$\theta^{II} = \frac{\partial c}{\partial r}\frac{1}{S} = \frac{\Delta \$\text{Pay}}{\Delta \$\text{Firm Value}} \tag{22}$$

$$\theta^{III} = \frac{\partial c}{\partial r} = \frac{\Delta \$\text{Pay}}{\Delta \ln \text{Firm Value}}. \tag{23}$$

where $r = V/S - 1$ is the firm's stock market return. Section 3.3 will predict how the three incentive measures scale with firm size and Section 3.4 will test these predictions. These tests will shed light on whether utility and production functions are additive or multiplicative, and thus the optimal measure of incentives.

We now solve for the second stage of Grossman and Hart (1983), i.e. the optimal effort level, returning to the case of additive preferences. With unlimited liability, the principal can always adjust fixed pay ϕ so that the participation constraint (16) binds. Thus, his expected pay is $\mathrm{E}[c(V)] = w + g(a^*)$, just as in the first-best, and so the principal's objective function remains (14). As a result, she implements the first-best effort level, defined by (15). Using (15) and (18), the optimal contract satisfies

$$\mathrm{E}[c'(V)\, b(S)] = b(S). \tag{24}$$

With a linear contract, this yields $\theta = 1$ and so the optimal contract is given by

$$c(V) = \phi + V, \text{ where} \tag{25}$$

$$\phi = w + g(a^*) - S - b(S)\, a^*. \tag{26}$$

The principal effectively "sells" the firm V to the agent for an up-front fee of $-\phi$, chosen so that the participation constraint (16) binds. Since the agent benefits one-for-one from any increase in firm value, he fully internalizes the benefits of effort and the first-best effort level a^*_{FB} is achieved. The level of incentives is "one size fits all": regardless of the cost or utility function, we have $\theta = 1$. There are no interesting comparative statics on how the contract differs between firms.

In the above framework, the effort level a^*_{FB} is chosen endogenously and so the principal implements whatever effort level is implied by $\theta = 1$. One simple way to obtain meaningful contracts that do differ across firms is to consider a binary effort decision, $a \in \{\underline{a}, \overline{a}\}$, where the principal implements $\overline{a}$, as in Holmstrom and Tirole (1997), Edmans, Gabaix, and Landier (2009), Biais et al. (2010), and the textbook of Tirole (2005). A similar specification is a continuous but bounded action space, $a \in [\underline{a}, \overline{a}]$, where again the principal wishes to implement $\overline{a}$. The upper bound reflects the fact that there may be a limit to the number of actions that a CEO can take to increase firm value. The high effort level $\overline{a}$ represents full productive efficiency, rather than working 24 hours a day. In a cash flow diversion model, full productive efficiency corresponds to zero stealing; in a project selection model, it corresponds to taking all positive-NPV projects while rejecting negative-NPV ones; in an effort model, it corresponds to the CEO not deliberately refraining from an action that will improve firm value because he prefers to shirk. Then, from equations (19) and (20), the optimal incentive level is $\theta b(S) = g'(\overline{a})$ if utility is additive and $\frac{\theta b(S)}{w} = g'(\overline{a})$ if utility is multiplicative.[12] Thus, the optimal level of incentives (\$-\$, \$-%, or %-% depending on the model specification) is increasing in the cost of effort $g'(\overline{a})$. Incentives are higher in firms with greater agency problems, rather than one-size fits all.

The first-best is still achieved in the fixed-action setting. In reality, the first-best cannot be achieved for two reasons. First, the agent may be subject to limited liability ($c(V) \geq 0$). Under contract (25), the agent will receive a negative payoff if V is sufficiently low, violating limited liability. Put differently, the agent may not have enough cash to buy the firm. Second, he may be risk-averse and demand a premium for bearing the risk associated with firm value V. We explore these two departures in turn and show that they both lead to non-trivial contracts.

Innes (1990) studies the case of limited liability and risk neutrality. The optimal contract is no longer linear and so we return to a general contract $c(V)$. He considers two versions of the model. In the first, the only restriction on the contract is limited liability on both the principal and agent, $0 \leq c(V) \leq V$. To keep the proof simple, we normalize w and set $g(0)$ to 0, although these assumptions are not necessary. Denote by $f(V, a)$ the probability density function of $V \in [0, \bar{V}]$ conditional on effort a and assume that it satisfies the monotone likelihood ratio property ("MLRP"), i.e.

$$\frac{f_a(V, a)}{f(V, a)}$$

is strictly increasing in V.

[12] When a is a boundary action, the IC becomes an inequality and a continuum of contracts will implement $a = \overline{a}$. We choose the contract that involves the minimum amount of incentives, as this is optimal for any non-zero level of risk aversion, and so the IC continues to bind.

The principal's problem is given by

$$\max_{c(\cdot)} \int_0^{\bar{V}} (V - c(V)) f(V, a^*) dV$$

$$\text{s.t.} \quad \int_0^{\bar{V}} c(V) f(V, a^*) dV - g(a^*) \geq w \tag{27}$$

$$\int_0^{\bar{V}} c(V) f_a(V, a^*) dV = g'(a^*) \tag{28}$$

$$0 \leq c(V) \leq V. \tag{29}$$

Note that for all contracts $c(\cdot)$ satisfying the IC (28), we have

$$\int_0^{\bar{V}} c(V) f(V, a^*) dV - g(a^*) \geq \int_0^{\bar{V}} c(V) f(V, 0) dV - g(0) = \int_0^{\bar{V}} c(V) f(V, 0) dV \geq 0,$$

where the first inequality arises because a^* maximizes the agent's utility if the IC (28) is satisfied, and the final inequality is due to the agent's limited liability, i.e. $c(V) \geq 0$. Thus, the IC (28) implies the IR (27) and so we can ignore the latter. We thus have the following Lagrangian:

$$L = \int_0^{\bar{V}} (V - c(V)) f(V, a^*) dV + \lambda \left(\int_0^{\bar{V}} c(V) f_a(V, a^*) dV - g'(a^*) \right),$$

which can be rewritten as

$$L = \int_0^{\bar{V}} c(V) f(V, a^*) \left(-1 + \lambda \frac{f_a(V, a^*)}{f(V, a^*)} \right) dV + \int_0^{\bar{V}} V f(V, a^*) dV - \lambda g'(a^*).$$

Pointwise optimization with respect $c(V)$, subject to the limited liability constraint (29), yields the following contract

$$c(V) = \begin{cases} 0 \text{ if } \frac{f_a(V,a^*)}{f(V,a^*)} < \frac{1}{\lambda} \\ V \text{ if } \frac{f_a(V,a^*)}{f(V,a^*)} \geq \frac{1}{\lambda} \end{cases}. \tag{30}$$

Due to MLRP, $f_a(V, a^*)/f(V, a^*)$ is strictly increasing. Thus, there exists an $\widehat{X}$ such that

$$c(V) = \begin{cases} 0 \text{ if } V < \widehat{X} \\ V \text{ if } V \geq \widehat{X} \end{cases}, \tag{31}$$

where $\widehat{X}$ is the largest X that satisfies the IC (28), which can be rewritten:

$$\int_X^{\bar{V}} V f_a(V, a^*) dV = g'(a^*).$$

Contract (31) is a "live-or-die" contract: the agent receives the entire firm value V if it exceeds a threshold $\widehat{X}$, and zero otherwise. The intuition is that the tails of the distribution are most informative about whether the agent has exerted effort. Thus, the optimal contract punishes the agent as much as possible for left-tail realizations of V, and rewards him as much as possible for right-tail realizations of V. With limited liability on the agent, he can receive no less than 0 for low outputs; with limited liability on the principal, she can pay no more than the entire firm value V for high outputs.

A potentially unrealistic feature of contract (31) is that it is discontinuous: when V rises from $\widehat{X}-\varepsilon$ to $\widehat{X}$, the principal's payoff falls from $\widehat{X}-\varepsilon$ to 0. Thus, the principal may wish to exercise her control rights on the firm to "burn" output from $\widehat{X}$ to $\widehat{X}-\varepsilon$ to increase her payoff. Alternatively, since the agent's pay rises more than one-for-one around this threshold, he may wish to borrow on his own account to increase output from $\widehat{X}-\varepsilon$ to $\widehat{X}$ because the gain in his payoff will exceed the amount he must repay. To deter both actions, the second version of the Innes (1990) model also assumes a monotonicity constraint: the principal's payoff cannot fall with firm value ($V-c(V)$ is nondecreasing in V), and so the agent's pay cannot increase more than one-for-one with firm value. Following similar steps to above, the optimal contract is very similar except that at the new cutoff $\widetilde{X}<\widehat{X}$, the contract jumps from 0 not to V, but only to $V-\widetilde{X}$, since this is the highest payoff that does not violate the monotonicity constraint. This yields the following contract:

$$c(V)=\begin{cases} 0 \text{ if } V<\widetilde{X} \\ V-\widetilde{X} \text{ if } V\geq\widetilde{X} \end{cases}, \tag{32}$$

where $\widetilde{X}$ is again the largest X that satisfies the IC (28), which can be rewritten

$$\int_{X}^{\bar{V}}(V-X)\,f_a(V,a^*)\,dV=g'(a^*).$$

Contract (32) is a standard call option, where the agent receives zero if V falls below a threshold X, and the residual $V-\widetilde{X}$ otherwise. The intuition is similar to contract (30): for low output ($V<\widetilde{X}$), the agent receives the lowest possible payoff (0); for high output ($V>\widetilde{X}$), he gains one-for-one with any increase in V which is the maximum possible gain without violating monotonicity.

Contract (32) implies not only that the CEO should be paid exclusively with options, but also that his wealth-performance sensitivity is 1 (for $V>\widetilde{X}$) – i.e. he gains dollar-for-dollar with any increase in firm value above $\widetilde{X}$.[13] Thus, the only source of variation between CEOs is the strike price $\widetilde{X}$. It is easy to show that, when the marginal cost of effort $c'(a^*)$ rises, the strike price falls in order to increase the delta of the option and maintain the agent's incentives.

[13] If there are x existing shares outstanding and the CEO is given options on y shares, his share of firm value is $\frac{y}{x+y}$ if he exercises all his options. Thus, strictly speaking, he must be given infinite options to obtain a wealth-performance sensitivity of 1.

Hence, even though the optimal contract is no longer trivial, this model does not capture much of the cross-sectional variation in real-life CEO contracts.

In reality, while CEOs are often paid with options in practice, they also receive salary, bonuses, and stock. Moreover, they often have very few shares compared to the number of shares outstanding, meaning that they gain far less than dollar-for-dollar with any increase in firm value. This wealth-performance sensitivity differs widely across firms, which the above model does not capture. We now incorporate risk aversion, which leads to the optimal sensitivity being below 1 and differing across CEOs. In addition to the strength of incentives, these models will also derive predictions for the optimal shape of contracts – whether they should be linear or convex, and thus whether they should comprise stock or options.

3.2. Risk-Averse Agent

3.2.1. Holmstrom-Milgrom Framework

Returning to the case of unlimited liability, another route to a meaningful contract is to have a risk-averse agent. Under the general utility function (10), and returning to general (rather than linear) contracts, the agent's first-order condition is given by:

$$\mathrm{E}\left[u'\left(\cdot\right)\left(v'\left(c\right)c'\left(V\right)b\left(S\right)-g'\left(a^{*}\right)\right)\right]=0. \tag{33}$$

Even assuming a given implemented action a^*, the contracting problem remains difficult because equation (33) only requires the contract to satisfy the agent's incentive constraint *on average*. Even in the simplest case in which u is linear, the agent's *average* expected marginal benefit from effort, $\mathrm{E}\left[v'\left(c\right)c'\left(V\right)b\left(S\right)\right]$, must equal the (known) marginal cost of effort, $g'\left(a^*\right)$. There are many potential contracts that will satisfy the incentive constraint on average, and so the problem is complex because the principal must solve for the one contract out of this continuum that minimizes the expected wage. (The problem is more complex if u is non-linear).

Holmstrom and Milgrom (1987, "HM") showed that the contracting problem becomes substantially simpler if four assumptions are made. First, the agent exhibits exponential utility, so $u\left(x\right)=-e^{-\eta x}$, where η is the coefficient of absolute risk aversion. Second, the cost of effort is pecuniary, so $v\left(c\right)=c$. Third, the noise ε is Normal, i.e. $\varepsilon\sim N\left(0,\sigma^2\right)$. Fourth, they consider a multi-period model where the agent chooses his effort every instant in continuous time. Under these assumptions, HM show that the optimal contract is linear, i.e. $c=\phi+\theta V$, and that the problem is equivalent to a single-period static problem. The intuition is that a linear contract subjects the agent to a constant incentive pressure irrespective of the history of past performance. This result suggests that incentives should be implemented purely with stock, and not non-linear instruments such as options.

The principal's problem becomes:

$$\max \mathrm{E}\left[V-c\right] \tag{34}$$

$$\text{s.t. } \mathrm{E}\left[-e^{-\eta\left[c-\frac{1}{2}ga^2\right]}\right] \geq -e^{-\eta w} \tag{35}$$

$$a \in \arg\max_{\widehat{a}} \mathrm{E}\left[-e^{-\eta\left[c-\frac{1}{2}g\widehat{a}^2\right]}\right]. \tag{36}$$

Substituting for $c=\phi+\theta V$ and $V=S+b\left(S\right)a+\varepsilon$, the agent's objective function simplifies to:

$$-e^{-\eta\widehat{c}(a)}, \tag{37}$$

where $\widehat{c}\left(a\right)=\phi+\theta\left(S+b\left(S\right)a\right)-\frac{1}{2}ga^2-\frac{\eta}{2}\theta^2\sigma^2$ is his utility from the contract. It comprises of the expected wage $\phi+\theta\left(S+b\left(S\right)a\right)$, minus the cost of effort $\frac{1}{2}ga^2$, minus the risk premium $\frac{\eta}{2}\theta^2\sigma^2$ that the agent requires. This risk premium is increasing in the agent's risk aversion η, risk σ^2, and incentives θ. The agent maximizes (37) by selecting

$$a^*=\frac{\theta b\left(S\right)}{g}. \tag{38}$$

His effort choice is independent of risk σ^2 and risk aversion η, since noise is additive. It is also independent of the fixed wage ϕ, since exponential utility removes wealth effects. Thus, ϕ can be adjusted to satisfy the agent's participation constraint without affecting his incentives.

Plugging (38) into the principal's objective function (34) and setting the participation constraint (35) to bind, the optimal level of incentives is

$$\theta=\frac{1}{1+g\eta\left(\frac{\sigma}{b(S)}\right)^2}. \tag{39}$$

Optimal incentives θ are a trade-off between two forces. A sharper contract increases effort $a^*=\frac{\theta b(S)}{g}$ and thus firm value, but also increases disutility $\frac{1}{2}ga^2$ and the risk premium $\frac{\eta}{2}\theta^2\sigma^2$. Thus, θ is decreasing in risk aversion η and risk σ^2 as these augment the risk premium required. The effect of the cost of effort g is more nuanced. On the one hand, fixing a^*, the required incentives are $\theta=\frac{a^*g}{b(S)}$ and is increasing in g. On the other hand, when effort is costlier to implement (g is higher), the optimal effort level a^* is lower. The second effect dominates: when effort is costlier, an increase in θ leads to a smaller rise in effort, and so the optimal θ falls. (Since the benefit of effort b has the opposite effect of the cost of effort g, we discuss only the latter throughout).

To find fixed pay ϕ, we set the participation constraint to bind ($\widehat{c}\left(a\right)=w$). This yields

$$\phi=w-\theta S-\frac{1}{2}\frac{\left(\theta b\left(S\right)\right)^2}{g}+\frac{\eta}{2}\theta^2\sigma^2.$$

The comparative statics for ϕ are ambiguous (see Appendix A). On the one hand, a higher cost of effort g, higher risk aversion η, and higher risk σ^2 increase the required fixed pay ϕ as a compensating differential (i.e. to ensure the IR remains satisfied). On the other hand, these changes also reduce the optimal level of incentives (from (39)), which lowers the risk premium.

The HM framework is attractive for a number of reasons. First, it derives (rather than assumes) a linear contract as being optimal. Second, it solves for not only the optimal contract to implement a given effort level, but also the optimal effort level, i.e. both stages of Grossman and Hart (1983). Solving for the optimal effort level is valuable not so much because empiricists test the model's predictions for the effort level (which is hard to observe), but – as will be made clear shortly – endogenizing the effort level leads to different predictions for the contract (which is observed). Third, the fixed salary ϕ does not affect the agent's effort choice. Thus, changes in reservation utility can be simply met by varying ϕ, without changing incentives.

However, HM stressed that a number of assumptions were necessary for their linearity result: exponential utility, a pecuniary cost of effort, Normal noise, and continuous time. Hellwig and Schmidt (2002) show that linearity continues to hold in discrete time under two additional assumptions: the principal does not observe the time path of profits (only the total profit in the final period), and the agent can destroy profits before he reports them to the principal. In Appendix B we discuss the role played by the first three assumptions.

3.2.2. Alternatives to Holmstrom-Milgrom

The HM model has proven extremely influential due to its tractability. Given the benefits of tractability, researchers have attempted to achieve tractability in other settings. We explore these alternative models here.

In HM, the effort level $a = \frac{\theta b(S)}{g}$ is chosen endogenously. As described in Section 3.1, an alternative specification is for the principal to implement a fixed target action $\overline{a}$. The optimal contract is now $\theta b(S) = \frac{g\overline{a}}{b}$, which leads to very different empirical implications. Now, the level of incentives $\theta b(S)$ (or $\frac{\theta b(S)}{w}$ with multiplicative utility) arises from the desire to induce effort $\overline{a}$, and not any trade-off with disutility or risk. Thus, only the first effect of g exists – a higher cost of effort raises the incentives required to induce $\overline{a}$ – and so incentives are increasing in g, in contrast to HM. It is also increasing in the target effort $\overline{a}$, but independent of η and σ^2, since the contract is not determined by any trade-off with these parameters. Thus, if the fixed action model accurately represents reality, it has the attractive practical implication that the contract does not depend on the agent's risk aversion, which is typically hard to observe. It thus offers a potential explanation for why real-world contracts do not seem as complicated and contingent on as many details of the environment as standard contract theories would suggest. For example, Section 3.4 shows that there is no systematic relationship between incentives and risk; the textbook of Bolton and Dewatripont (2005, p158) notes that "what is surprising is the relative simplicity of observed managerial compensation packages given the complexity of the incentive problem". These details do not matter because the contract is determined by the need to induce effort $\overline{a}$, rather than a trade-off with risk. In addition, we now have unambiguous predictions for how increases in risk σ^2 and risk aversion η affect the level of pay. There is now only the direct effect, that pay rises as a compensating differential, but no indirect effect because these parameters do not affect the optimal effort level.

Whether the endogenous or fixed action model is more realistic depends on the setting. In many cases, the endogenous action case is more accurate as principals choose to implement less-than-full effort to save on wages. For example, a factory boss may only require a production operative to work an eight-hour day, to avoid paying overtime. However, for CEOs, a fixed action may be more appropriate. Edmans and Gabaix (2011b) show that, if CEO effort has a multiplicative effect on firm value, implementing full productive efficiency $\overline{a}$ is optimal if the firm is large enough. (The result also holds for any increasing function $b(S)$). The benefits of effort are a function of firm size; the cost of effort (a higher wage to compensate for risk and disutility) is a function of the CEO's reservation wage w. Thus, if S is sufficiently large compared to w, the benefits of effort dominate the trade-off and it is optimal to induce full productive efficiency regardless of g, η or σ^2. For example, in a \$10bn firm, if implementing effort level $\overline{a} - \xi$ rather than $\overline{a}$ reduces firm value by only 0.1%, this translates into \$10m. If the CEO salary is \$10m, even if salary can be reduced by 50% by allowing the CEO to exert only $\overline{a} - \xi$, implementing $\overline{a}$ remains optimal. Indeed, the structural estimation of Margiotta and Miller (2000) finds that the costs of inducing effort are substantially less than the benefits. The fixed action model more likely applies to CEOs than rank-and-file employees, who have a limited effect on firm value.

The overall point that we would like to stress is not that one model is superior to the other. Different models apply to different scenarios. Rather, we wish to highlight how a contracting model's empirical implications hinge critically on the assumptions – whether we specify multiplicative versus additive production or preference functions, or a fixed versus continuous implemented action. Sometimes, researchers may assume a binary action space or additive functions out of convenience, but this modeling choice can lead to vastly different predictions.

The framework of Edmans and Gabaix (2011b, "EG") provides another way to obtain tractable contracts, without the need to assume exponential utility, a pecuniary cost of effort, or Normal noise. It considers the implementation of a given effort level a^*, i.e. the first stage of Grossman and Hart (1983), and thus is particularly applicable to CEOs where $a^* = \overline{a}$ if the firm is large. EG specify the noise ε as being realized before, rather than after the action a is taken. This timing is also featured in models in which the agent observes total cash flow before deciding how much to divert (e.g., Lacker and Weinberg (1989), Biais, Mariotti, Plantin, and Rochet (2007), DeMarzo and Fishman (2007)), and in which he observes the "state of nature" before choosing effort (Harris and Raviv (1979), Sappington (1983), Baker (1992), and Prendergast (2002)). Note that this timing assumption does not render the CEO immune to risk, because noise is unknown when he signs his contract. EG also show that the contract retains the same form in continuous time, where noise and effort occur simultaneously. This consistency suggest that, if underlying reality is continuous time, it is best approximated in discrete time by modeling noise before effort.

The timing assumption allows for significant tractability. Since the noise is known when the agent takes his action, we can remove the expectation from his objective function (10) to yield:

$$u\left(v\left(c\left(S+b\left(S\right)a+\varepsilon\right)\right)-g\left(a\right)\right). \tag{40}$$

In turn, $u(\cdot)$ also drops out. The specific form of u is irrelevant – since it is monotonic, it is maximized by maximizing its argument. This yields the first-order condition:

$$v'\left(c\left(S+b\left(S\right)a^*+\varepsilon\right)\right)c'(S+b\left(S\right)a^*+\varepsilon)b\left(S\right)=g'\left(a^*\right). \tag{41}$$

This first-order condition must hold for every possible ε, i.e., *state-by-state*, rather than simply on average. This pins down the slope of the contract: for all ε, the agent must receive a marginal felicity of $g'(a^*)$ for a marginal increase in V. Thus, for all V, the contract must satisfy:

$$v'\left(c\left(V\right)\right)c'\left(V\right)b\left(S\right)=g'\left(a^*\right).$$

Integrating over 0 to V, we obtain in felicity units:

$$v\left(c\left(V\right)\right)=\frac{g'\left(a^*\right)}{b\left(S\right)}V+k,$$

for an integration constant k. This yields, in dollar terms,

$$c\left(V\right)=v^{-1}\left(\frac{g'\left(a^*\right)}{b\left(S\right)}V+k\right) \tag{42}$$

The constant k is chosen to make the participation constraint bind, i.e.

$$\mathrm{E}\left[u\left(\frac{g'\left(a\right)}{b(S)}V+k-g\left(a^*\right)\right)\right]=w. \tag{43}$$

There is a unique optimal contract. The slope is chosen so that the incentive constraint (41) holds state-by-state, and the scalar k is chosen so that the participation constraint binds.

Equation (42) shows that the optimal contract is typically non-linear. Even though the noise is known when the agent takes his action, it is not irrelevant because it has the potential to undo the agent's incentives. If ε is high, V and thus c will already be high; a high reservation wage w increases the required constant k and thus c, and so has the same effect. If the agent exhibits diminishing marginal felicity (i.e., v is concave), he has lower incentives to exert effort. Put differently, the agent does not face *risk* (as ε is known) but *distortion* (as ε affects his effort incentives). HM assume that the cost of effort is in financial terms so that – like the benefit of effort – it also declines with ε, and so incentives are unchanged with a linear contract. EG instead address distortion by the shape of the contract: it is convex, via the v^{-1} transformation. If noise is high, the contract gives a greater number of dollars for each incremental unit of firm value ($c'(V)$), to offset the lower marginal felicity of each dollar ($v'(c)$). Therefore, the marginal felicity from effort remains $v'(c)c'(V)\,b(S) = g'(a^*)$, and incentives are preserved regardless of w or ε. Allowing for convex contracts removes the need to assume a pecuniary cost of effort. In contrast, if v is convex, the contract is concave.

The contract is linear in two special cases. The first is a pecuniary cost of effort, as in HM: when $v(c) = c$, we have $c(V) = \frac{g'(a^*)}{b(S)}V + k$. With an additive production function ($\gamma = 0$), the CEO's dollar incentives are linear in the firm's dollar value V; with a multiplicative production function ($\gamma = 1$), they are linear in the firm's percentage return $\frac{V}{S}$. The former result echoes Lacker and Weinberg (1989), who also feature a pecuniary cost of effort and an additive production function. They show that the optimal contract to deter all cash flow diversion (the analogy of $a^* = \overline{a}$) is piecewise linear. The second case is $v(c) = \ln c$, i.e. multiplicative preferences. The contract is $\ln c(V) = \frac{g'(a^*)}{b(S)}V + k$, and so log pay is linear in V ($\frac{V}{S}$) with an additive (multiplicative) production function. In both cases, the framework delivers linear contracts without requiring exponential utility or Normal noise. More broadly, the framework allows for contracts that are convex and concave, rather than purely linear as in HM – thus, tractability can be achieved without linearity – and shows what determines the optimal curvature or linearity of the contract: the form of $v(\cdot)$.

Equation (42) also clarifies the parameters that do and do not matter for the contract's functional form. It depends only on the felicity function v and the cost of effort g. The functional form is independent of the utility function u, the reservation utility w, and the distribution of the noise ε, i.e. the contract can be written without reference to these parameters. These parameters will still affect the contract's slope via their impact on the scalar k. However, the contract's slope as well as its functional form are independent of u, w, and ε in the cases of $v(c) = c$ and $v(c) = \ln c$, where it depends only on $g'(a^*)$. This "detail-independence" contrasts with standard agency models where the contract depends on many specific features of the setting.[14]

The above framework allows for tractable contracts with fewer restrictions on the utility function, cost of effort, and noise distribution, as well as non-linear contracts. This tractability allows it to be used in dynamic models with private saving (Edmans, Gabaix, Sadzik, and Sannikov (2012)) and assignment models with moral hazard under risk aversion (Edmans and Gabaix (2011a)). However, it has a number of disadvantages. It requires the assumption that noise precedes the action; while applicable in some settings (e.g. a cash flow diversion model), it may not apply to others. It also assumes a fixed implemented action a^*, which again may only apply in some settings (e.g. a CEO of a large firm). The goal of this article is to provide a range of modeling frameworks, each of which may be applicable under different conditions.

[14] Chassang (2013) derives contracts that are relatively independent of the environment (i.e. the probability space), in a risk-neutral setting.

We have so far considered two models that yield tractable contracts at the cost of some assumptions. Other papers do not aim to achieve a tractable analytical solution, but instead to calibrate the optimal contract, and so use fewer assumptions. Perhaps the most commonly used framework for calibration involves constant relative risk aversion ("CRRA") utility and lognormal firm value, studied by Lambert, Larcker, and Verrecchia (1991), Hall and Murphy (2000, 2002), Hall and Knox (2004), and others. We present here the version calibrated in Dittmann and Maug (2007). End-of-period firm value is given by

$$V_T = V_0(a) \exp \left[\left(R - \frac{\sigma^2}{2} \right) T + \varepsilon \sigma \sqrt{T} \right],$$

where $\varepsilon \sim N(0,1)$ and V_0 satisfies $V_0' > 0$ and $V_0'' < 0$. They assume that a contract is composed of salary ϕ, θ shares, and ψ options (as a fraction of shares outstanding) with strike price X and maturity T.[15] Both shares and options are paid out at the end of the period; salary ϕ is paid out at the start. The CEO begins with non-firm wealth W_0, which is invested at the risk-free rate R. His end-of-period wealth is then given by

$$c_T = (\phi + W_0)e^{RT} + \theta V_T + \psi \max\{V_T - X, 0\}.$$

In the general utility function (10), we have $u(x) = x$ and $v(c_T) = \frac{c_T^{1-\zeta}}{1-\zeta}$, where ζ is the parameter of relative risk aversion. Thus, the CEO's preferences are given by

$$U(c_T, a) = \frac{c_T^{1-\zeta}}{1-\zeta} - g(a).$$

Assuming risk-neutral pricing, the CEO's end-of-period pay (change in wealth) is given by

$$\pi_T = \phi e^{RT} + \theta V_T + \psi \max\{V_T - X, 0\}.$$

with expected present value

$$\pi_0 = E[e^{-RT}\pi_T] = \phi + \theta V_0 + \psi BS,$$

where BS denotes the Black-Scholes value of the option. They solve for the first stage of Grossman and Hart (1983), in which the principal wishes to implement action a^*, and so her

[15] In an additional analysis, Dittmann and Maug (2007) also solve for the optimal unrestricted contract.

problem is given by

$$\min_{\phi,\theta,\psi} \pi_0 = \phi + \theta V_0 + \psi BS$$
$$\text{s.t. } E[U(W_T, a^*)] \geq \bar{U},$$
$$a^* = \arg\max_{a\in[0,\infty)} E[U(W_T, a)]$$
$$\phi + W_0 \geq 0, \qquad 0 \leq \theta \leq 1, \qquad \psi \geq 0$$

Dittmann and Maug (2007) calibrate this model to a sample of 598 US CEOs. In particular, they study whether it is more efficient to incentivize the CEO with stock or options. Since options are riskier, \$1 of options is worth less to the CEO than \$1 of stock, rendering them less effective in meeting the CEO's participation constraint. On the other hand, \$1 of options provides greater incentives than \$1 of stock, rendering them more effective in meeting his incentive constraint. They find that the first effect is dominant, suggesting that the optimal contract should involve only stock and not options. This prediction is shared with Holmstrom and Milgrom (1987) who predict linear contracts, although in a different setting. Moreover, when they drop the restriction that the contract must be piecewise linear (i.e. consist of salary, stock, and options), they find that the optimal nonlinear contract is concave.

In contrast to both frameworks, option compensation is widespread in the U.S. One interpretation, consistent with Bebchuk and Fried (2004), is that the use of options indicates rent extraction: since options did not have to be expensed until 2006, they constitute "stealth compensation" not noticed by shareholders. Indeed, Murphy (2013) discusses how accounting rules and disclosure requirements affect the mix of stock vs. options. For example, the use of options fell substantially after FAS 123(R) mandated that their value be expensed; previously, options did not need to be expensed if they were at-the-money. However, Dittmann, Maug, and Spalt (2010) show that options can be rationalized if the CEO is loss-averse: since options provide downside protection, they are particularly valuable to a loss-averse agent. Moreover, as we will discuss in Section 3.6, if the agent chooses firm risk in addition to effort, options may be useful to induce him to take value-adding risky projects.

3.3. Incentives in Market Equilibrium

Section 3 has thus far taken the reservation wage w as given. We now endogenize w using the assignment model of Gabaix and Landier (2008) to study how CEO incentives vary across firms in market equilibrium. We use the Edmans, Gabaix, and Landier (2009) framework of a risk-neutral CEO, multiplicative preferences and a fixed target action, as in Section 3.1, with $a^* = \overline{a}$. We will show that even this simple model leads to predictions consistent with empirical findings. (Edmans and Gabaix (2011a) extend the model to risk aversion.)

From (20), we have $\theta = \frac{\Lambda w}{S}$ where $\Lambda = g'(a^*)$. The fixed salary ϕ is chosen so that the IR binds, i.e. $\phi = w - \theta S = w(1-\Lambda)$. Thus, the CEO in firm n is given a fixed salary ϕ^*, and θS worth of shares, with:

$$\theta_n S_n = w(n)\Lambda, \quad (44)$$

$$\phi_n = w(n)(1-\Lambda), \quad (45)$$

where $w(n)$ is given by equation (8) from Gabaix and Landier (2008). Thus, a fraction Λ of the equilibrium wage is paid in equity, and the remainder is paid in cash.

We can now solve for the three incentive measures in equations (21)-(23) in terms of model primitives:

$$\theta^I = \Lambda \propto S^0 \quad (46)$$

$$\theta^{II} = \Lambda \frac{w}{S} \propto S^{\rho-1} \quad (47)$$

$$\theta^{III} = \Lambda w \propto S^{\rho}, \quad (48)$$

Equation (20) earlier suggested that, in a multiplicative model, the optimal incentive measure is θ^I (%-% incentives) since it affects the implemented effort level. Equations (21)-(23) illustrate a related advantage: in a multiplicative model, θ^I is independent of firm size and thus comparable across firms of different size. Intuitively, since effort has a percentage effect on both firm value and CEO utility, it is %-% incentives that are relevant. Comparability across firms of different size is useful to study which firms are incentivized more or less than their peers. For example, a passive investor who believes that incentives are not fully priced in the market may wish to invest in a stock with high CEO incentives; an activist investor may wish to target a firm with low incentives. However, if the CEO of a large firm has \$2m of equity and the CEO of a smaller firm has \$1m of equity, we cannot immediately conclude which CEO is better incentivized as dollar equity holdings should optimally increase with firm size. Relatedly, comparability is valuable for boards or compensation consultants undertaking benchmarking analyses. [16]

While %-% incentives should be independent of size, with $\rho = \gamma - \beta/\alpha = 1/3$ as in Gabaix and Landier (2008), \$-\$ incentives should have a firm-size elasticity of $\rho - 1 = -2/3$. If effort has a multiplicative effect on firm value, it has a higher dollar effect in a larger firm, and so a lower equity stake is needed to induce effort. In addition, \$-% incentives should have an elasticity of $\rho = 1/3$. Larger firms hire more talented CEOs who command higher wages. Since the benefits of shirking are higher, given multiplicative preferences, a higher dollar equity stake is needed to induce effort. In Section 3.4, we will compare these predictions to the data.

[16] By analogy, fund managers are compared according to their risk-adjusted percentage returns, rather than dollar returns, as the former is comparable across funds of different size (assuming constant returns to scale).

Turning to the strength of incentives, the \$-\$ incentives measured by Jensen and Murphy (1990) are given by $\theta^{II} = \theta^{I} \frac{w}{S}$. Since firm size S is substantially larger than the CEO's wage w, \$-\$ incentives should be low. Because firms are so large, the dollar benefits of effort are much greater than the disutility cost to the CEO, and so only a small equity stake is needed to induce effort. Another strand of research justifies low \$-\$ incentives by pointing out the disadvantages of strong incentives. Lambert, Larcker, and Verrecchia (1991) show that a high equity stake may induce the CEO to take inefficiently low risk. Benmelech, Kandel, and Veronesi (2010) assume that equity incentives vest in the short-term, since long-term incentives expose the CEO to risks outside his control. Then, the CEO may conceal information that his investment opportunities have declined to keep the current stock price high, even though disclosing such information will allow him to efficiently disinvest. In a similar vein, Peng and Roell (2008, 2014) and Goldman and Slezak (2006) demonstrate that high-powered incentives, that vest in the short term, can encourage the manager to expend firm resources to manipulate the stock price upwards.

3.4. Empirical Analyses

We now turn to tests of the empirical predictions of these models. The first set of tests study the level of incentives. Motivated by traditional additive models, Jensen and Murphy (1990) estimate \$-\$ incentives and showed that the CEO loses only \$3.25 for every \$1,000 loss in firm value, an effective equity stake of only 0.325%. They interpreted this stake as too low to be reconciled with optimal contracting, and thus concluded that CEOs are "paid like bureaucrats". However, such a conclusion hinges critically on whether we believe CEO effort has additive or multiplicative effects on firm value and CEO utility. \$-\$ incentives are the relevant measure only in an additive model. As discussed above, in a multiplicative model, %-% incentives are relevant and \$-\$ incentives should optimally be low.

Separately, theory predicts that incentives should be $\frac{\overline{a}g}{b(S)}$ or $\frac{1}{1+g\eta\left(\frac{\sigma}{b(S)}\right)^2}$, but parameters such as the cost of effort g are difficult to quantify. Thus, it is difficult to evaluate whether quantitative findings on the level of incentives are consistent with efficiency. Haubrich's (1994) calibration suggests that the seemingly low incentives found by Jensen and Murphy (1990) can be optimal if the CEO is sufficiently risk-averse, but attaches wide confidence intervals to his conclusion given the difficulties in calibration. The structural estimation of Margiotta and Miller (2000) also finds that low incentives are sufficient to induce effort given the multiplicative effect of effort on firm value.

Given the difficulties of quantifying parameters such as g to calculate the optimal level of incentives, incentive theories are typically tested instead in terms of their cross-sectional predictions – whether they vary with parameters such as S, g, η and σ^2 as predicted. Note that it is important for empirical tests to study the precise measure of incentives predicted by the theory. For example, if the theory is a multiplicative model that predicts how the dollar equity stake θS varies with g, η, and σ^2, studying the percentage equity stake θ will not be a precise test of the model as these parameters may vary with firm size S. In addition, equation (39) implies that with a multiplicative production function ($b(S) = S$), the relevant measure of risk is $\frac{\sigma}{S}$, the volatility of the firm's percentage returns; with an additive production function ($\gamma = 0$) it is the σ, the volatility of the firm's dollar returns.

Starting with size, Jensen and Murphy (1990) found that \$-\$ incentives are even lower in large firms, perhaps because governance is particularly weak in these firms. As discussed above, Edmans, Gabaix, and Landier (2009) show that, under a multiplicative model, CEO effort has a larger dollar effect in a bigger firm, and so a smaller equity stake is required to induce effort. They quantitatively predict a firm-size elasticity of $-2/3$, consistent with their empirical finding of -0.60. Similarly, they find that \$-\$ incentives are independent of firm size and \$-% incentives have a size-elasticity of $1/3$, both as predicted. Thus, a model with multiplicative utility and production functions quantitatively explains the size-scalings of incentives. While these results are consistent with incentives being set optimally and the true model indeed being multiplicative, they could also be consistent with a non-multiplicative model and suboptimal incentive setting.

We now turn to HM's prediction that incentives θ are decreasing in risk σ. While Lambert and Larcker (1987), Aggarwal and Samwick (1999), and Jin (2002) indeed find a negative relationship, Demsetz and Lehn (1985), Core and Guay (1999), Oyer and Schaefer (2005), and Coles, Daniel, and Naveen (2006) document a positive relationship, and Garen (1994), Yermack (1995), Bushman, Indjejikian, and Smith (1996), Ittner, Larcker, and Rajan (1997), Conyon and Murphy (1999), Edmans, Gabaix, and Landier (2009), and Cheng, Hong, and Scheinkman (2015) show either no relationship or mixed results. Aggarwal and Samwick (1999) and Jin (2002) study the volatility of dollar returns, and the other papers study percentage returns. Thus, the empirical evidence points to a weak relationship between risk and incentives. The fixed action model provides a potential explanation: risk is second-order compared to the benefits of effort – it is incentive considerations, not risk considerations, that affect the slope of the contract.[17]

The prediction that θ is decreasing in risk aversion η is harder to test as risk aversion is unobservable. Becker (2006) uses data on CEO wealth, available in Sweden, as a (negative) proxy for risk aversion under the assumption of decreasing absolute risk aversion. As predicted, he finds that wealth is positively related to both \$-\$ and %-% incentives.[18] In addition, wealth can affect incentives through channels other than risk aversion. In the EG model, where the

[17] Prendergast (2002) provides another explanation for the weak relationship between risk and incentives. When uncertainty is low, principals assign tasks to agents and directly monitor them. When uncertainty is high, they delegate tasks to agents and incentivize them through output. His model applies principally to rank-and-file employees, since day-to-day monitoring of the CEO by directors is more limited.

[18] While HM assume CARA utility and so risk aversion is independent of wealth, the model of Sannikov (2008), analyzed in Section 4.2, generally predicts that incentives fall with risk aversion by the same intuition as in HM. His model allows for general utility functions, and thus absolute risk aversion to be decreasing in wealth.

contract is not driven by a trade-off with risk aversion, the CEO's outside option w may include consuming his existing wealth. Higher wealth increases w and thus the constant k (equation (43)), which in turn augments incentives (equation (42)). Intuitively, if the CEO is wealthier, his marginal utility from money is lower, and so greater incentives are required to induce him to work.

The theories also derive predictions for expected pay $E[c]$, often referred to as the level of pay. As discussed, firm risk and disutility have an ambiguous effect on the level of pay in the HM model, but increase it in the fixed action model due to the required compensating differential. Garen (1994) shows that pay is insignificantly increasing in firm risk as measured by dollar volatility, and insignificantly decreasing in percentage volatility. Cheng, Hong, and Scheinkman (2015) find a significant positive relationship with percentage volatility for financial firms. Gayle and Miller (2009) show theoretically and empirically that larger firms are more complex to manage, and so CEOs require greater pay in return. In addition, greater agency problems in large firms necessitate higher equity incentives and thus more pay as a risk premium. Conyon, Core, and Guay (2011) and Fernandes, Ferreira, Matos, and Murphy (2013) compare CEO pay in the U.S. to the rest of the world, and show that the pay premium to U.S. CEOs can be explained by the greater risk that they bear, rather than rent extraction. The structural estimation of Gayle, Golan, and Miller (2015) finds that the risk premium can explain over 80% of the pay differential between small and large firms. It arises both because large firms require greater incentives to address moral hazard, and also because stock returns are a poorer signal of effort in large firms.

3.5. Multiple Signals

The analysis has thus far studied the sensitivity of the manager's pay to the performance of his own firm. Here, we study the extent to which it should depend on other signals, such as industry and market conditions. We first demonstrate the Holmstrom (1979) informativeness principle. This principle considers the case in which, in addition to firm value V, the principal has access to an additional contractible signal z (such as the performance of peers) and studies the extent to which CEO pay c should depend on z. The joint density function is given by $f(V, z, a)$, and the principal's problem is:

$$\max_{c(\cdot,\cdot)} \int_z \int_0^{\bar{V}} (V - c(V,z)) f(V,z,a^*) dV dz \tag{49}$$

$$\text{s.t.} \int_z \int_0^{\bar{V}} u(c(V,z)) f(V,z,a^*) dV dz \geq g(a^*) \tag{50}$$

$$\int_z \int_0^{\bar{V}} u(c(V,z)) f_a(V,z,a^*) dV dz = g'(a^*) \tag{51}$$

Denote by λ and μ the Lagrange multipliers for the IR (50) and IC (51), respectively. Pointwise optimization yields the following condition for the optimal contract $c(V, z)$:

$$\frac{1}{u'(c(V, z))} = \lambda + \mu \frac{f_a(V, z, a^*)}{f(V, z, a^*)}$$

for all (V, z). Hence, assuming that $\mu \neq 0$ (i.e. the IC binds), the contract $c(V, z)$ is not a function of z if and only if the likelihood ratio $f_a(V, z, a^*)/f(V, z, a^*)$ does not depend on z, i.e.

$$\frac{f_a(V, z, a^*)}{f(V, z, a^*)} = h(V, a^*). \tag{52}$$

for some function h. Condition (52) holds if and only if V is a sufficient statistic for $\{V, z\}$ with respect to $a = a^*$.[19] Thus, any signal z, no matter how noisy, that provides information incremental to V on the agent's effort choice, should be included in the contract.

The most common application of the informativeness principle to CEO pay is relative performance evaluation ("RPE"). Specifically, peer performance is informative about the degree to which high firm value V is due to high effort or good luck. Conventional wisdom is that RPE is very rarely used in reality. Aggarwal and Samwick (1999) and Murphy (1999) show that CEO pay is determined by absolute, rather than relative performance, and Jenter and Kanaan (2015) find an absence of RPE in CEO firing decisions. However, Gong, Li, and Shin (2011) argue that these conclusions arise from identifying RPE based on an implicit approach – assuming a peer group (e.g. one based on industry and/or size) and relevant performance measures, and studying whether those performance measures for that peer group affect CEO pay. These may assumptions lead to measurement error that biases downwards the estimated use of RPE. Gong et al. study the explicit use of RPE, based on the disclosure of peer firms and performance measures mandated by the SEC in 2006. They find that 25% of S&P 1500 firms explicitly use RPE. Still, the practice is not widespread, and Bebchuk and Fried (2004) argue that the rarity of RPE is a key piece of evidence in support of the rent extraction view. Bertrand and Mullainathan (2001) demonstrate that pay-for-luck is strongest in poorly-governed firms, consistent with the view that it is inefficient.

There are several potential justifications for the rarity of RPE. The first is to consider it in context: there is little indexation of wage contracts in the economy in general, even though it would not only allow for more precise signals to satisfy the incentive constraint, but also aid in satisfying the participation constraint. For example, indexing wages to variables such as inflation, growth, or the business cycle would help ensure that workers' reservation utility is always met. The lack of indexation for CEOs is consistent with its absence for rank-and-file

[19] Condition (52) is required to be satisfied only for effort level $a = a^*$; it may be that the likelihood ratio depends on z for a different $a \neq a^*$. We thus say that (52) holds if and only if V is a sufficient statistic for $\{V, z\}$ with respect to $a = a^*$. Holmstrom (1979) assumes that condition (52) is satisfied either for all a or no a, and so he shows that the contract is a function of z if and only if V is a sufficient statistic for $\{V, z\}$ with respect to a (rather than $a = a^*$).

employees. Since these employees are unlikely to have influence over the contract designers, it is unlikely that the lack of indexation represents rent extraction. The deeper reasons for the general absence of indexation are still unclear, but a potential explanation is preference for simplicity and flexibility in contracting.

Second, Dittmann, Maug, and Spalt (2013) show that the indexation of options can destroy incentives. Since an indexed option is in the money only if the stock price rises high enough to outperform the benchmark, indexation is tantamount to increasing the strike price of an option and reducing the drift rate of the underlying asset. Both effects reduce the option's delta and thus his incentives. To preserve incentives, additional equity must be given, and their calibration shows that full indexation of all options would increase compensation costs by 50% on average. If firms choose the optimal proportion of options to index, average compensation costs would only fall by 2.3%, and 75% of firms would choose zero indexation. They show that indexing stock also has little benefit. Chaigneau, Edmans, and Gottlieb (2015a) similarly show that indexation can reduce incentives in the Innes (1990) framework of risk neutrality and limited liability, which allows for an optimal contracting approach. Third, Oyer (2004) shows that, if equity is forfeited upon departure, it induces the agent to stay with the firm. Since non-indexed equity is more valuable in high market conditions, when the outside option is also higher, its retention power increases precisely when retention concerns are greatest. Fourth, Gopalan, Milbourn, and Song (2010) argue that not indexing an executive to industry performance induces him to choose the firm's industry exposure optimally.

In addition to signals about peer performance, the informativeness principle implies that any informative signal, no matter how noisy, should be in the optimal contract. In reality, in addition to firm value, CEO pay may depend on accounting performance measures (such as sales growth, return-on-assets, and earnings per share growth) through their impact on discretionary bonuses, as well as subjective evaluations by principals (e.g. Cornelli, Kominek, and Ljungqvist (2013)). However, CEO pay does not appear to depend on non-accounting performance measures such as surveys on intangible assets (e.g. customer satisfaction, brand strength, and employee engagement) and the number of patent citations. These all potentially provide information over and above that contained in the stock price, since the stock market does not immediately capitalize intangibles.

However, as stressed by Holmstrom (1979), the informativeness principle as derived assuming no constraints on the contract.[20] In reality, contracting constraints do exist, such as limited liability. Chaigneau, Edmans, and Gottlieb (2015b) show that the informativeness principle may not hold under limited liability. In the standard Innes (1990) framework, the agent receives zero below a threshold and gains one-for-one above the threshold, which is the maximum possible without violating the agent's monotonicity constraint (if imposed) or principal's lim-

[20] This observation also explains why Dittmann, Maug, and Spalt (2013) find that indexation may destroy value. In reality, contracts consist of a restricted set of securities – salary, bonus, shares, and options – but the optimal contract is very far from piecewise linear (see, e.g., Dittmann and Maug (2007)).

ited liability (if monotonicity is not imposed). Since constraints on the contract are binding almost everywhere, the principal's ability to use the signal is severely restricted. If low firm value V is accompanied by a low signal z, she cannot punish the agent further without violating limited liability as he is already receiving zero. Her only degree of freedom is on the level of the threshold, and so a signal is only valuable if it affects the optimal cutoff.

This result suggests that the common practice of paying agents for luck is not necessarily suboptimal. If a firm suffers a catastrophe, the manager is typically paid zero, regardless of whether it was down to bad luck (e.g. poor industry performance) or shirking. In reality, instances of "pay for luck" typically concern very good or very bad outcomes – for example, Bertrand and Mullainathan (2001) consider how CEO pay varies with spikes and troughs in the oil price, and Jenter and Kanaan (2014) find that peer-group performance does not affect CEO firing decisions – but additional signals are only valuable for moderate outcomes. However, it does not explain why signals on non-accounting performance are not used in normal times, i.e. outside of firing considerations.

3.6. Risk-Taking

Thus far, the CEO takes an action that changes the firm's expected value, but has no direct effect on its risk. In Smith and Stulz (1985), the agent takes a single action that reduces risk via hedging. If the agent is risk averse, he will engage in excessive hedging; in an CEO context, this corresponds to turning down positive-NPV risky projects. They show how options address this issue, since their convexity counterbalances the concavity of the agent's utility function. Dittmann and Yu (2011) calibrate a model where the CEO chooses both effort and risk, and show that it can explain the mix of stock and options found empirically. However, Carpenter (2000) and Ross (2004) show theoretically that options may not increase the manager's risk-taking incentives: while an option has "vega" (positive sensitivity to volatility), it also has "delta" (positive sensitivity to firm value). This, a risk-averse manager may wish to reduce volatility in the value of the firm and thus his options. Shue and Townsend (2013) evaluate this theoretical debate empirically by showing that exogenous increases in options, resulting from their multi-year grant cycles, lead to an increase in risk-taking.

The above models consider "good" risk-taking that improves firm value. However, the CEO may also have incentives to engage in "bad" risk-taking that reduces firm value. In particular, in a levered firm, an equity-aligned manager may undertake a project even if it is negative-NPV, because shareholders benefit from the upside but have limited downside risk due to limited liability (Jensen and Meckling (1976)). Anticipating this, creditors will demand a high cost of debt and/or tight covenants, to the detriment of shareholders.

Edmans and Liu (2010) show that a potential solution to such risk-shifting is to compensate the CEO with debt as well as equity. (Such debt is referred to as "inside" debt, as it is owned by the manager rather than outside creditors.) Previously proposed remedies for risk-shifting include bonuses for achieving solvency, or salaries and private benefits that are forfeited in bankruptcy (e.g. Brander and Poitevin (1992)). These instruments are sensitive to the incidence of bankruptcy, but if bankruptcy occurs, they pay zero regardless of liquidation value. In contrast, inside debt yields a positive payoff in bankruptcy, proportional to the recovery value. Thus it renders the manager sensitive to firm value in bankruptcy, and not just the incidence of bankruptcy – exactly as desired by creditors – and thus reduce the cost of raising debt, to the benefit of shareholders. Indeed, recent empirical studies have shown that CEOs hold a substantial amount of inside debt through defined benefit pensions and deferred compensation. These are unsecured obligations which yield an equal claim with other creditors in bankruptcy, and thus constitute inside debt. For example, Sundaram and Yermack (2007) show that GE's Jack Welch had over $100 million of inside debt when he retired in 2001.

Since traditional contracting theories typically advocate only the use of equity, and disclosure of pensions and especially deferred compensation was limited prior to a 2007 SEC disclosure reform, Bebchuk and Fried (2004) argue that inside debt constitutes "stealth compensation through retirement benefits". However, the risk deterrence story suggests that inside debt can be consistent with optimal contracting. The disclosure of significant inside debt positions following the SEC reform led to an increase in bond prices (Wei and Yermack (2011)), and is associated with lower bond yields and fewer covenants (Anantharaman, Fang, and Gong (2014), using personal state income taxes as an instrument for inside debt). Debt-aligned executives manage the firm more conservatively as measured by the firm's lower distance to default (Sundaram and Yermack (2007)), and lower stock return volatility, R&D expenditures and financial leverage (Cassell, Huang, Sanchez, and Stuart (2012), also using the income tax instrument). Indeed, the alignment of executives with debt has gathered pace in the recent crisis. In 2010, American International Group tied 80% of highly paid employees' pay to the price of its bonds, and 20% to the price of its stock, and UBS and Credit Suisse have since started paying bonuses in bonds. The Liikanen Report of the European Commission and the Federal Reserve have advocated debt-like compensation to curb excessive risk-taking. However, even if the above studies can be interpreted as showing causal effects of inside debt on risk-taking and borrowing conditions, they do not study whether shareholders benefit overall, nor whether alternative solutions to risk-shifting would be superior.

In Smith and Stulz (1985), the firm is unlevered so there are no risk-shifting concerns; the contract contains options but no debt. In Edmans and Liu (2010), the CEO is risk-neutral so there is no problem of inducing him to take "good" risk; the contract contains debt, but not options. For future research, it would be interesting to incorporate both leverage and risk aversion into a model of both effort and risk-taking, to study the optimal mix of salary, stock, options, and debt.

4. Dynamic Incentives

This section analyzes dynamic models of moral hazard. In reality, CEOs are employed for several years. A dynamic setting leads to additional questions, such as how to spread the rewards for good performance over time, how the level and sensitivity of pay vary over time, and when the CEO quits or is fired. We start in Section 4.1 with a tractable discrete-time model that yields closed-form solutions, at the cost of some assumptions. In Section 4.2 we move to a continuous-time model which typically yields numerical solutions, but allows for departures and terminations.

4.1. Dynamic Incentives: A Simple Discrete Time Model

We present here the discrete-time version of the Edmans, Gabaix, Sadzik, and Sannikov (2012) model, which uses the EG framework to yield tractable solutions. In every period t, the CEO takes first observes noise ε_t and then takes action a_t, which affects terminal (period-T) firm value as follows

$$V_T = Se^{\sum_{s=1}^{T}(a_s+\varepsilon_s)}.$$

We assume that a signal about it, $V_t = Se^{\sum_{s=1}^{t}(a_s+\varepsilon_s)}$, is contractible. The incremental information contained in V_t over and above that contained in V_{t-1} can be summarized by the stock "return"

$$r_t = \ln V_t - \ln V_{t-1} = a_t + \varepsilon_t$$

where, as in EG, the CEO observes the noise ε_t before he takes his action a_t. Also in every period t, the principal pays the CEO $y_t\left(r_1, ..., r_t\right)$ which may depend on the entire history of returns. The agent consumes c_t and saves $\left(y_t - c_t\right)$ (which may be positive or negative) at the continuously compounded risk-free rate R. He lives for T periods and retires after period $L \leq T$.[21]

The agent's lifetime utility is:

$$U = \sum_{t=1}^{T} e^{-\delta t}\mathsf{u}\left(c_t, a_t\right), \qquad \mathsf{u}\left(c, a\right) = \ln c - g(a) \tag{53}$$

where δ is the agent's discount rate, i.e. his impatience. His per-period utility function $\mathsf{u}\left(c, a\right) = \ln c - g\left(a\right)$ corresponds to (10) with $v\left(c\right) = \ln c$ and $u\left(x\right) = x$, i.e. multiplicative preferences. The agent's reservation utility is w.

The principal is risk-neutral and so her objective function is expected discounted terminal

[21] In the model, the principal replaces the CEO with a new one and continues to contract optimally, but this assumption can easily be weakened.

firm value, minus expected pay:

$$\max_{(a_t,t=1,\ldots L),(y_t,t=1,\ldots T)} E\left[e^{-RT}V_T - \sum_{t=1}^{T} e^{-Rt}y_t\right]. \tag{54}$$

She wishes to implement a target action sequence (a_t^*).

There are two constraints to consider. The first is the effort constraint ("EF"), which ensures that the CEO does not wish to deviate from (a_t^*). We consider a local deviation in the action a_t after history $(r_1, \ldots r_{t-1}, \varepsilon_t)$. The effect on CEO utility should be zero:

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

Since $\partial r_t/\partial a_t = 1$ and $\partial U/\partial a_t = e^{-\delta t}\mathsf{u}_a(c_t, a_t)$, the EF constraint (evaluated at $a_t = a_t^*$) is

$$\begin{aligned} \text{EF} : E_t\left[\frac{\partial U}{\partial r_t}\right] &= e^{-\delta t}\mathsf{u}_a(c_t, a_t^*) \text{ if } a_t^* \in (0, \overline{a}) \\ E_t\left[\frac{\partial U}{\partial r_t}\right] &\geq e^{-\delta t}\mathsf{u}_a(c_t, a_t^*) \text{ if } a_t^* = \overline{a} \end{aligned} \tag{55}$$

for $t \leq L$.

The second constraint is the private savings constraint ("PS"), which ensures that the CEO consumes his income in period t, i.e. $c_t = y_t$, so that he has no incentive to save privately. If the CEO saves a small amount d_t in period t and invests it until $t+1$, his utility increases to the leading order by $-E_t\left[\frac{\partial U}{\partial c_t}\right]d_t + E_t\left[\frac{\partial U}{\partial c_{t+1}}\right]e^R d_t$. To deter private saving or borrowing, this change should be zero to the leading order,

$$\text{PS} : 1 = E_t\left[e^{R-\delta}\frac{\mathsf{u}_c(c_{t+1}, a_{t+1})}{\mathsf{u}_c(c_t, a_t)}\right], \tag{56}$$

that is, the consumption Euler equation should hold. If the CEO cannot engage in private saving (e.g. because the principal can observe saving), then instead the inverse Euler equation ("IEE") holds, as is standard in agency problems with additively separable utility (e.g. Rogerson (1985) and Farhi and Werning (2012)). This is given by

$$\text{IEE} : 1 = E_t\left[\frac{1}{e^{R-\delta}}\frac{\mathsf{u}_c(c_t, a_t)}{\mathsf{u}_c(c_{t+1}, a_{t+1})}\right] \tag{57}$$

We next present the solution (a heuristic proof is in Appendix A). For simplicity, we assume a constant target action ($a_t^* = a^*$, which may correspond to full productive efficiency $\overline{a}$). The contract is given by:

$$\ln c_t = \ln c_0 + \sum_{s=1}^{t}(\theta_s r_s + k_s), \tag{58}$$

where θ_s and k_s are constants. The sensitivity θ_s is given by

$$\theta_s = \begin{cases} \frac{g'(a^*)}{1+e^{\delta}+...+e^{\delta(T-s)}} \text{ for } s \leq L, \\ 0 \text{ for } s > L. \end{cases} \tag{59}$$

If private saving is impossible, the constant k_s ensures that the IEE (57) holds:

$$k_s = R - \ln E\left[e_s^{\theta_s(a^*+\varepsilon)}\right]. \tag{60}$$

If private saving is possible, k_s ensures that the PS constraint (61) holds:

$$k_s = R + \ln E\left[e_s^{-\theta_s(a^*+\varepsilon)}\right] \tag{61}$$

The closed-form solutions allow transparent economic implications. Equation (58) shows that time-t income should be linked to the return not only in period t, but also in all previous periods. Therefore, increases in r_t boost log pay in the current and all future periods equally. Since the CEO is risk-averse, it is efficient to reward for good performance over the future to achieve consumption smoothing: the "deferred reward" principle. This result was first derived by Lambert (1983) and Rogerson (1985), who consider a two-period model where the agent only chooses effort.

We now consider how contract sensitivity changes over time. Equation (59) shows that, in an infinite-horizon model ($T \to \infty$), the sensitivity is constant and given by

$$\theta_t = \theta = \left(1 - e^{-\delta}\right) g'\left(a^*\right). \tag{62}$$

This is intuitive: the contract must be sufficiently sharp to compensate for the disutility of effort, which is constant. The sensitivity to the current-period return is decreasing in the discount rate – if the CEO is more impatient (higher δ), it is necessary to reward him today more than in the future.

If T is finite, equation (59) shows that θ_t is increasing over time: the "increasing incentives principle". When there are fewer periods over which to spread the reward for effort, the current-period reward ($\partial \mathsf{u}_t/\partial a_t = \theta_t$) must increase to keep the lifetime increase in utility $\partial U/\partial a_t$ constant. Other moral hazard models predict increasing incentives through different channels. In contrast, Gibbons and Murphy (1992) generate an increasing current sensitivity because the lifetime increase in utility $\partial U/\partial a_t$ rises over time to offset falling career concerns. In Garrett and Pavan (2015), the current sensitivity rises over time because $\partial U/\partial a_t$ increases to minimize the agent's informational rents. Here, $\partial U/\partial a_t$ is constant since we have no adverse selection or career concerns; instead, the increase in $\partial \mathsf{u}_t/\partial a_t$ stems from the reduction in consumption smoothing possibilities as the CEO approaches retirement.

While θ_t depends on the model horizon, it is independent of whether private saving is possible – this possibility only affects k_t. Since private saving does not affect the agent's action and thus firm returns, the sensitivity of pay to returns is unchanged. From (58), the possibility of private saving alters the time trend in the level of pay. The log expected growth rate in pay is $\ln E\left[c_t/c_{t-1}\right] = k_t + \ln E\left[e^{\theta_t r_t}\right]$.

If private saving is impossible, substituting for k_t using (60) yields

$$\ln E\left[\frac{c_t}{c_{t-1}}\right] = R - \delta,$$

which is constant over time. If and only if the CEO is more patient than the aggregate economy ($\delta < R$), then the growth rate is positive, as is intuitive. If private saving is possible, (61) yields

$$\ln E\left[\frac{c_t}{c_{t-1}}\right] = R - \delta + \ln E[e^{-\theta_t r_t}] + \ln E[e^{\theta_t r_t}].$$

In the limit of small time intervals (or, equivalently, in the limit of small variance of noise σ^2), this yields

$$\ln E\left[\frac{c_t}{c_{t-1}}\right] = R - \delta + \theta_t^2 \sigma_t^2.$$

Thus, the growth rate of consumption is always higher when private saving is possible. This faster upward trend means that the contract effectively saves for the agent, removing the need for him to do so himself. This result is consistent with He (2012), who finds that the optimal contract under private saving involves a wage pattern that is non-decreasing over time.[22] The model thus predicts a positive relationship between pay and tenure, consistent with the common practice of seniority-based pay. Moreover, the growth rate depends on the risk to which the CEO is exposed, which is in turn driven by his incentives θ and firm volatility σ. This is intuitive: greater risk increases the CEO's motive to engage in precautionary saving (since, with CRRA utility, $u'''(c) > 0$), and so a rapidly-rising level of pay is necessary to remove the need for him to save privately. Furthermore, in a finite-horizon model, θ_t is increasing over time and so the growth rate of consumption rises with tenure, that is, pay accelerates over time.

To illustrate the economic forces behind the contract, we now present a simple numerical example with $T = 3$, $L = 3$, $\delta = 0$, $a_t^* = a^*$, and $g'(a^*) = 1$. From (59), the contract is:

$$\begin{aligned}
\ln c_1 &= \frac{r_1}{3} + \kappa_1, \\
\ln c_2 &= \frac{r_1}{3} + \frac{r_2}{2} + \kappa_2, \\
\ln c_3 &= \frac{r_1}{3} + \frac{r_2}{2} + \frac{r_3}{1} + \kappa_3,
\end{aligned}$$

[22] Lazear (1979) has a back-loaded wage pattern for incentive, rather than private saving considerations (the agent is risk-neutral in his model). Since the agent wishes to ensure he receives the high future payments, he induces effort to avoid being fired. Similarly, in Yang (2009), a back-loaded wage pattern induces agents to work to avoid the firm being shut down.

where $\kappa_t = \sum_{s=1}^{t} k_s$. An increase in r_1 leads to a permanent increase in log consumption – it rises by $\frac{r_1}{3}$ in all future periods. In addition, the sensitivity $\partial \mathsf{u}_t / \partial a_t$ increases over time, from 1/3 to 1/2 to 1/1. The total lifetime reward for effort $\partial U_t / \partial a_t$ is a constant 1 in all periods.

We now consider $T = 5$, so that the CEO lives after retirement. The contract is now

$$
\begin{aligned}
\ln c_1 &= \frac{r_1}{5} + \kappa_1, \\
\ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2, \\
\ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3, \\
\ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_4, \\
\ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_5.
\end{aligned} \tag{63}
$$

Since the CEO takes no action from $t = 4$, his pay does not depend on r_4 or r_5. However, it depends on r_1, r_2, and r_3 as his earlier efforts affect his wealth, from which he consumes.

We finally extend the basic model to allow the agent to engage in short-termism, and study how this possibility affects the optimal contract. Short-termism is broadly defined to encompass any action that increases current returns at the expense of future returns – scrapping positive-NPV investments (see, for example, Stein (1988)) or taking negative-NPV projects that generate an immediate return but weaken long-run value (such as sub-prime lending), earnings management, and accounting manipulation.

At time t, in addition into an effort and savings decision, the manager can also take a myopic action $m_{t,i}$ that increases the current return to $r_t' = r_t + M_i(m_{t,i})$ where M_i is a concave function. The CEO also chooses a "release lag" i, which is the number of periods before the negative consequences of myopia become evident. The maximum possible release lag is $H \leq T - L$. Myopia at t with release lag i reduces the return at $t+i$ to $r_{t+i}' = r_{t+i} - m_{t,i}$, and leaves other returns unchanged ($r_{t+s}' = r_{t+s}$ for $s \neq 0, i$). Let $\mathcal{M}_i = M_i'(0) \in [0,1)$ denote the marginal inefficiency of manipulation at release lag i.

If the firm is sufficiently large, the principal will wish to implement zero manipulation, i.e. $m_{t,i} = 0 \; \forall \; t$. If the agent engages in a small myopic action $m_{t,i} \geq 0$ at time t, his utility changes to the leading order by

$$E_t \left[\frac{\partial U}{\partial r_t} \right] \mathcal{M} m_{t,i} - E_t \left[\frac{\partial U}{\partial r_{t+i}} \right] m_{t,i}.$$

This should be weakly negative for all small $m_{t,i} \geq 0$. Hence, we obtain an additional No Manipulation ("NM") constraint:

$$\text{NM} : E_t \left[\frac{\partial U}{\partial r_t} \right] \mathcal{M}_i \leq E_t \left[\frac{\partial U}{\partial r_{t+i}} \right] \tag{64}$$

for $t \leq L$. The optimal contract is now as above, with the additional constraint (64).

We apply this constraint to the 5-period model of equation (63), with $H = 1$. The optimal contract now changes to:

$$\begin{aligned} \ln c_1 &= \frac{r_1}{5} + \kappa_1, \\ \ln c_2 &= \frac{r_1}{5} + \frac{r_2}{4} + \kappa_2, \\ \ln c_3 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \kappa_3, \\ \ln c_4 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{\mathcal{M}_1 r_4}{2} + \kappa_4, \\ \ln c_5 &= \frac{r_1}{5} + \frac{r_2}{4} + \frac{r_3}{3} + \frac{\mathcal{M}_1 r_4}{2} + \kappa_5. \end{aligned} \tag{65}$$

Even though the CEO retires at the end of $t = 3$, his income depends on r_4, otherwise he would have an incentive to boost r_3 at the expense of r_4. Thus, the CEO should retain equity in the firm even after retirement. For a general maximum release lag of H, the CEO should be sensitive to firm returns until period $L + H$, i.e. retain equity in the firm for H years after retirement. This result formalizes the verbal argument of Bebchuk and Fried (2004), who advocate escrowing the CEO's equity to deter him from inflating the stock price before retirement and then cashing out. For example, Angelo Mozilo, the former CEO of Countrywide Financial, made \$129 million from stock sales in the 12 months prior to the start of the subprime crisis. Indeed, in the aftermath of the crisis, banks such as Goldman Sachs and UBS have been increasing vesting horizons. An alternative remedy proposed is to use clawbacks, i.e. pay executives bonuses for good short-term performance, but rescind them if the performance ends up being reversed in the long-term. However, the legality of such clawbacks is unclear, and even if legally possible, they may be costly to implement. Lengthening the vesting period of equity so that rewards are not paid out prematurely in the first place may be a superior solution.

The sensitivity to r_4 depends on the efficiency of earnings inflation $\mathcal{M}_1$; in the extreme, if $\mathcal{M}_1 = 0$, myopia is impossible and so there is no need to expose the CEO to returns after retirement. The contract is unchanged for $t \leq 3$, that is, for the periods in which the CEO works. Even under the original contract, there is no incentive to inflate earnings at $t = 1$ or $t = 2$ because there is no discounting, and so the negative effect of myopia on future returns reduces the CEO's lifetime utility by more than the positive effect on current returns increases it. With discounting, incentives increase even faster over time than in the absence of a myopia problem. The higher sensitivity to future returns ensures that myopia causes the CEO to lose enough in the future to counterbalance the effect of discounting.

The contracts in (63) and (65) can be implemented in a simple manner. Each year, the manager's annual pay is escrowed into an "Incentive Account", a proportion θ_t of which is invested in stock and the remainder in cash, so his %-% incentives equal θ_t given by (59). If the stock price declines, so that the fraction of stock falls below θ_t, cash in the account is used to buy stock to replenish his incentives. Every year, a fraction of the account vests and is paid to the manager, but the remainder remains escrowed to deter myopia. Zhu (2014) shows that "bonus banks", introduced in practice by the consulting firm Stern Stewart, are a similar way to deter myopia. Bonuses for short-term performance are deposited into the "bonus bank", rather than immediately paid to the manager, and only a fraction is paid each period. Poor performance in one period, which may be caused by a myopic action in a previous period, wipes out previously accrued bonuses.

The advantage of the above framework is that it yields closed-form solutions that make the economic intuition transparent. However, it comes at the cost of a number of assumptions. First, as in the EG model, it assumes a fixed target action and that the noise precedes the action, which may be reasonable in some settings but not others. Second, it assumes a fixed retirement date T and does not allow for quits or firings beforehand, which is an important limitation in a CEO setting.

4.2. Dynamic Incentives in Continuous Time

This section presents the continuous-time model of Sannikov (2008) which allows for quits and firings, as well as the implemented effort level to be endogenized. At every instant t, the agent takes action a_t and consumes c_t; the framework rules out private saving so we do not distinguish between income and consumption. His expected lifetime utility at date t is

$$U_t = E_t \left[\int_t^\infty e^{-\delta(s-t)} \mathsf{u}\left(c_s, a_s\right) ds \right]. \tag{66}$$

It is the "promised utility" that the agent will obtain if he exerts the path of efforts recommended by the principal. This path of efforts is given by an adapted process (a_t^*).

The principal uses the same discount rate δ, and her expected utility is:

$$Q_t = E_t \left[\int e^{-\delta(s-t)} \left(dV_t - c_t\right) ds \right], \tag{67}$$

where dV_t is the evolution of firm value, assumed to be:

$$dV_t = a_t dt + \sigma dZ_t. \tag{68}$$

To derive the optimal contract, we first recall two basic lemmas from stochastic calculus, proven in Appendix A.

Martingale representation theorem. If $U_t = E_t \left[\int_t^\infty e^{-\delta(s-t)} K_s ds\right]$ for some adapted process K_t, then

$$dU_t = \left(-K_t + \delta U_t\right) dt + \xi_t dZ_t, \tag{69}$$

where ξ_t is some adapted process.

Hamilton-Jacobi Bellman ("HJB") equation. If

$$Q(x) = \sup_{(C_s)_{s \geq t}} E_t \left[\int_t^{\infty} e^{-\delta(s-t)} f(x_s, C_s)\, ds \mid x_t = x \right]$$

for some control variable C_t and stochastic process x_t with $dx_t = \mu(x_t, C_t)\, dt + \sigma(x_t, C_t)\, dZ_t$, then $Q(\cdot)$ satisfies:

$$0 = \sup_C f(x, C) - \delta Q(x) + Q'(x)\, \mu(x, C) + \frac{1}{2} Q''(x)\, \sigma^2(x, C) \tag{70}$$

Using the martingale representation theorem (69), the agent's utility process can be written:

$$dU_t = (-\mathsf{u}(c_t, a_t) + \delta U_t)\, dt + \beta_t \sigma dZ_t$$

for some process β_t that represents the sensitivity of the contract (equation (69) yields a process ξ_t, and we set $\beta_t = \xi_t / \sigma$). On the equilibrium path, $dV_t = a_t^* dt + \sigma dZ_t$, so the contract is:

$$dU_t = (-\mathsf{u}(c_t, a_t^*) + \delta U_t)\, dt + \beta_t (dV_t - a_t^* dt)\,. \tag{71}$$

The intuition is as follows. U_t represents an "account" which contains the lifetime utility promised to the agent. If the principal provides instantaneous utility to the agent, the account falls by $-\mathsf{u}(c_t, a_t^*)$ as less utility is owed in the future. If firm value is higher than expected ($dV_t - a_t^* dt > 0$), the account rises in value. Otherwise, it grows at a rate δ (the δU_t term).

We now move to the agent's IC. If the agent chooses action a_t, his utility is

$$\max_{a_t} \mathsf{u}(c_t, a_t)\, dt + dU_t(a_t)$$

i.e., he receives $\mathsf{u}(c_t, a_t)$ today, and will receive continuation utility of $dU_t(a_t)$ later. Given (68) and (71), we have

$$dU_t = \beta_t a_t dt + \text{terms independent of } a_t,$$

which yields the IC

$$a_t^* \in \arg\max_a \mathsf{u}(c_t, a_t) + \beta_t a_t.$$

The first-order condition yields:

$$\beta_t = -\mathsf{u}_a(c_t, a_t)\,, \tag{72}$$

i.e. incentives β_t are determined by c_t and a_t^*.

Turning to the principal's problem, the state variable for the HJB stochastic process is $x_t = U_t$, the agent's promised utility. The principal's value function is $Q(x_t)$. Using (72), the

HJB equation (70) is, for all x,

$$0 = \max_{c,a} a - c - \delta Q(x) + Q'(x)[-\mathsf{u}(c,a) + \delta x] + \frac{1}{2} Q''(x) \mathsf{u}_a(c,a)^2 \tag{73}$$

This gives an ordinary differential equation for $Q(x)$. From this, the optimal c and a are implicitly determined by:

$$-1 - Q'(x)\mathsf{u}_c(c,a) + \frac{1}{2} Q''(x) \partial_c \left(\mathsf{u}_a(c,a)^2\right) = 0 \tag{74}$$

$$1 - Q'(x)\mathsf{u}_a(c,a) + \frac{1}{2} Q''(x) \partial_a \left(\mathsf{u}_a(c,a)^2\right) = 0. \tag{75}$$

We now need to specify the boundary conditions. The agent's per-period reservation utility is w, and so he accepts the contract only if $U_t \leq w/\delta$. Let $Q_0(x) = -\mathsf{u}(\cdot,0)^{-1}(\delta w)/V + A$ denote the cost of providing this to agent when the agent exerts zero effort, where A is the present value of the principal's outside option, e.g. his surplus from hiring a new agent. The agent is employed if and only if his promised utility is $x \in [x_L, x_H]$, with the following matching and smooth pasting conditions:

$$Q(x) = Q_0(x), Q'(x) = Q_0'(x) \text{ for } x = x_L, x_H. \tag{76}$$

Hence, the problem is characterized by the ODE (73), x_L, and x_H. At x_L, the agent is terminated because of poor performance. At x_H, he is terminated because he has become too expensive to incentivize. Intuitively, as the agent's promised utility increases, his marginal utility of money falls, and so it is harder to incentivize him.

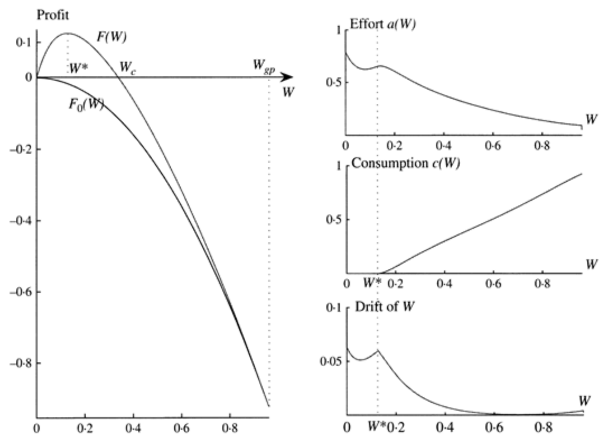
We have four unknowns, x_L, x_H, and the two degrees of freedom associated with a second-order ODE (given a value x_*, the ODE is described by two parameters, $Q(x_*)$ and $Q'(x_*)$). We also have four equations (76). Hence, the problem yields a solution.

To sum up, the principal solves the problem as follows. Given her value function Q, she finds the optimal action and consumption from the first-order conditions (74) and (75). These in turn determine the optimal incentives from (72), and the agent's utility is given by (71).

This problem is quite complex, and typically the solutions are numerical.[23] Here, we reproduce a result from Sannikov (2008):[24]

[23] One closed-form solution is the one in the Edmans, Gabaix, Sadzik, and Sannikov (2012) setup, with a constant a^* (e.g. $a^* = a_H$) and no outside opportunity ($w = -\infty$). With log utility ($\mathsf{u}(c,a) = \ln c - g(a)$) and $r = \rho$, the reader can verify that $c(x) = De^{rx}$, $Q(x) = Ae^{rx} + B$ for some constants A, B, D. With constant relative risk aversion ($\mathsf{u} = \left(ce^{-g(a)}\right)^{1-\gamma} / (1-\gamma)$), the solution has the form $c = Dx^{1/(1-\gamma)}, Q(x) = Ax^{1/(1-\gamma)} + B$. The resulting contracts are described in Section 4.1. Otherwise, the only known solutions are numerical.

[24] We are in the process of adapting this figure to change notation as follows: $W \to w, F(W) \to Q(x), W_{gp} \to x_H$.



Function F for $u(c) = \sqrt{c}$, $h(a) = 0{\cdot}5a^2 + 0.4a$, $r = 0{\cdot}1$ and $\sigma = 1$. Point W^* is the maximum of F

While the numerical results depend somewhat on the utility function, for the specification in Figure 1, when promised utility rises, consumption increases. As a result, effort falls: the marginal utility of additional consumption is low and so stronger monetary incentives are required to induce effort. Since the agent is now expensive to incentivize, the optimal effort level falls. Other variables have non-monotonic relationships with promised utility. One important open question would be to take these theoretical models to the data, and determine which functional form best describe the world. The absence of analytical solutions renders comparative statics relatively difficult to obtain. In many situations, greater risk aversion reduces incentives, for the same intuition as Holmstrom and Milgrom (1987) – incentives are more costly as they require the agent to be paid a greater risk premium.

Sannikov's (2008) methodology has since been used in executive compensation applications. He (2012) extends the model to allowing for private saving. In standard models without private saving (e.g. Rogerson (1985)), the optimal wage profile is front-loaded, but such a profile will induce the agent to engage in a joint deviation of shirking and saving. He shows that the wage profile is back-loaded, to deter such private saving. He also finds that pay does not fall upon poor performance but exhibits a permanent rise after a sufficiently good performance history. This downward rigidity is also predicted by Harris and Holmstrom (1982), but through a quite different channel. Their model features two-sided learning about the agent's ability rather than moral hazard. Downward rigidity in wages insures the agent against negative news about his ability, while wage rises after positive news ensure that he does not quit.[25]

[25] DeMarzo and Sannikov (2006) use the Sannikov (2008) framework to study optimal capital structure and show that it can be implemented with standard securities – a credit line, long-term debt, and outside equity. Since the agent always holds equity, the model focuses on financing rather than executive compensation.

4.3. Empirical Analyses

We now turn to tests of the empirical predictions of dynamic models. Lambert (1983), Rogerson (1985), and Edmans, Gabaix, Sadzik, and Sannikov (2012) predict the "deferred reward principle": firm performance should affect future as well as current pay due to consumption smoothing considerations. Indeed, Boschen and Smith (1995) show that firm performance has a much greater effect on the NPV of future pay than current pay. Gibbons and Murphy (1992) find support for the "increasing incentives principle", that incentives rise over time. This result is consistent with both consumption smoothing possibilities and career concerns falling with tenure. In addition to incentives, Murphy (1986) finds that pay increases over time, consistent with models which predict a backward-loaded wage pattern to remove incentives for private saving. However, to our knowledge, predictions that the growth rate in pay depends on the level of incentives θ and firm risk σ are as yet untested.

Turning to the predictions regarding termination, many models predict termination after poor performance, to deter shirking ex ante, e.g. DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), Biais, Mariotti, Plantin, and Rochet (2007), and Sannikov (2008).[26] In particular, the first three models feature limited liability, which reduces the principal's ability to punish poor performance financially, thus leading to a role for termination. In some cases, such as Sannikov (2008), termination also arises after very good performance as the agent becomes too expensive to incentivize. However, historically, dismissals have been relatively rare. Murphy and Zabojnik (2007) and Jensen and Murphy (2004) report turnover rates of 10% per year in the 1970s and 1980s, which increased only to 11% in the 1990s. Taylor's (2010) learning model shows that the low rate of dismissals can only be justified if the costs to shareholders of turnover exceed \$200 million. Turning to turnover-performance sensitivity, Coughlan and Schmidt (1985), Jensen and Murphy (1990), Warner, Watts, and Wruck (1988), and Weisbach (1988) find that turnover probability is decreasing in performance but the economic magnitude is small. Jensen and Murphy (1990) estimate that a CEO who performs in line with the market over 2 years has a 11.1% dismissal probability; underperforming by 50% in each year increases this probability only to 17.5%. Thus, even under the aggressive assumption that the CEO receives no severance package and is unable to find alternative employment until retirement, Jensen and Murphy (1990) estimate that incentives from dismissal are equivalent to an equity stake of 0.03%.

Moreover, incentives from dismissal are even lower if the CEO is granted severance pay. In contrast to most theories, which advocate that pay should be weakly monotonic in firm performance, CEOs are often given severance packages upon departure. Yermack (2006) finds a mean contracted severance pay of \$0.9 million, with a mean discretionary amount of \$4.5 million;

[26] Note that termination after poor performance is typically not subgame-perfect, so moral hazard models assume that the firm can commit to terminate the CEO. Learning models predict subgame-perfect termination after poor performance, as such performance signals low managerial quality (e.g. Jovanovic (1979), Taylor (2010), and Garrett and Pavan (2012)).

the respective maximums are \$36.1 and \$121.1 million, suggesting that these packages can be substantial. Their usage is especially prevalent among dismissed CEOs compared to those who voluntarily retire, and thus appears to reward CEOs for failure. However, a closer look at the data suggests that the vast majority of "severance pay" does not stem from compensation for loss of employment, but instead items such as unvested restricted shares, unexercised stock options, and accrued pension benefits, which were promised and contractually obligated to the CEO under any state of nature. For example, out of Henry McKinnell's much-criticized \$180m severance package from Pfizer, \$78m was deferred compensation (\$67m contributed plus \$11m interest), \$82m was the present value of his pension plan, and \$8m was from stock options. Thus, only an incremental \$11m was due to the loss of employment.[27] Furthermore, some theories predict that severance pay can be optimal. Almazan and Suarez (2003) show that it can induce the CEO to leave voluntarily when a more able replacement is available; Inderst and Mueller (2008) demonstrate that it can deter a CEO from entrenching himself by concealing negative information that would lead to his dismissal. One example is to induce the CEO to accept a takeover bid, which typically yields a substantial premium to shareholders. Manso (2011) shows that severance pay is valuable to induce the CEO to explore new technologies rather than merely exploit existing ones. In the aforementioned model of He (2012), severance pay leads to a backward-loaded wage pattern that is robust to private savings.

Recent studies of CEO turnover have uncovered higher rates. Kaplan and Minton (2011) aggregate both internal (board-driven) and external (through takeover and bankruptcy) turnover and find that total annual turnover was 17.4% over 1998-2005. A one standard-deviation fall in the industry-adjusted stock return is associated with a 3.4% increase in the likelihood of turnover. This figure is 2.1% for industry performance and 1.8% for the performance of the overall market. Jenter and Lewellen (2014) find a total turnover rate of 11.8% per year of which they estimate 4.1-4.5 percentage points (35-38% of the total) are performance-induced, i.e. would not have occurred had performance been good.

Thus, more recent evidence suggests that the threat of job loss from poor performance is significant. While these results support the prediction that firing occurs upon poor performance, they do not support the prediction that it is also prompted by good performance. In addition, moral hazard models typically do not yield quantitative predictions for what the rate of firing or the sensitivity of firing to performance should be, making it difficult to assess whether the observed findings are optimal.[28]

[27] We thank David Yermack for this example.

[28] Taylor (2010) derives quantitative predictions for the rate of firing as a function of the cost of turnover to shareholders, but in a learning model.

5. Open Questions

5.1. Apparent Inefficiencies in Executive Compensation

Thus far, we have argued that many features of executive compensation that are frequently criticized may yet be consistent with efficient contracting. Examples include the level of pay, low \$-\$ incentives, the negative relationship between incentives and firm size, the use of stock rather than options, the lack of relative performance evaluation, and the use of severance pay and inside debt. However, this empirical consistency does not prove that compensation practices are efficient. As discussed earlier, the positive correlation between pay and firm size is also consistent with rent extraction, and may arise because a third omitted variable drives both; similar concerns surround other empirical findings consistent with optimal contracting models. In addition, even accepting the empirical correlations discussed in this paper as supportive of optimal contracting in general, there remain several features of compensation that could be improved upon.

First, empirical studies uncover results for the average firm. However, even if compensation is efficient on average, there may still be several individual cases of rent extraction. For example, while the theories discussed in Section 4.3 may be able to justify the mean level of severance pay, their forces are unlikely to be strong enough to rationalize extreme realizations, such as the maximum discretionary award of \$121.1 million. In addition, while Dittmann, Maug, and Spalt (2013) find that indexation of options would not create value for the average firm, it would for 25% of firms and so the absence of indexation across all firms is difficult to reconcile with efficiency.

Second, some aspects of compensation are hidden from shareholders, which is difficult to reconcile with them being set in shareholders' interest. For example, Lie (2005) presents evidence that the positive stock returns after the disclosed grant dates of executive stock options, first documented by Yermack (1997), arises from backdating. Since options are typically granted at the money, the CEO – unbeknown to shareholders – chooses the grant date in retrospect, to coincide with days on which the stock price is low and thus justifying a low strike price. Similarly, recent corporate scandals such as Tyco uncovered executives extracting perks that were initially unknown to shareholders.

Third, current schemes often fail to keep pace with a firm's changing conditions. While Section 3.4 argues that the CEO's incentives are sufficient in normal times to induce effort, if a company encounters difficulties and its stock price falls, the delta of his options decline and so they lose much of their incentive effect. One remedy used in practice is the repricing of out-of-the-money options (Brenner, Sundaram, and Yermack (2000); Acharya, John, and Sundaram (2000)), but this is controversial as it appears to reward the CEO for failure. Even if the CEO is paid purely with stock (which always has a delta of 1), the problem continues to exist as long as the benefit of effort $b(S)$ is increasing in firm size. Intuitively, when firm size falls, the benefits from effort are lower and so additional equity is needed to induce a given level of effort. For example, if both the production and utility functions are multiplicative, the relevant measure is %-% incentives, the percentage of the CEO's wealth that is comprised of stock, and this measure falls when the stock price declines. If the utility function is additive, the relevant measure is $-% incentives, the dollar value of the CEO's equity, which also falls when the stock price declines. A simple way to replenish incentives is to increase (reduce) the portion of the CEO's salary that is given in equity (cash) after the stock price falls; indeed, Core and Guay (1999) show that firms use new equity grants to move executives towards their optimal incentive levels. Alternatively, the Incentive Account discussed in Section 4.1 involves rebalancing the amount of the CEO's escrowed equity and deferred cash to ensure he always has sufficient equity. Critically, in both cases, unlike the repricing of options, the CEO's additional equity is not given for free: it is paid for by a reduction in cash. Thus, the CEO is reincentivized without him being rewarded for failure.

Fourth, standard measures of CEO incentives, such as those considered in Section 3, only measure how the CEO's wealth is affected by changes to the current stock price. They do not consider the extent to which the CEO is also aligned with the long-term stock price, i.e. the horizon of incentives. The classic managerial myopia models of Stein (1988, 1989) show that short-term incentives can lead to myopic actions. In a corporate context, these actions can involve cutting R&D, reducing employee training, writing loans that may become delinquent in the future, or expending corporate resources on earnings management. Empirical studies of the horizon of incentives have been hindered by lack of data availability on the vesting period of an executive's equity, but recent studies suggest that horizons may affect behavior. Gopalan, Milbourn, Song, and Thakor (2014) use the recent change in disclosure requirements to pioneer a measure of the "duration" of incentives, analogous to the duration of a debt security. Shorter duration incentives are correlated with earnings management. Edmans, Fang, and Lewellen (2015) study the quantity of equity scheduled to vest in a given year, since this amount depends on equity grants made several years prior and is thus likely exogenous to current investment opportunities. They find that vesting equity is significantly negatively correlated with cuts in R&D, advertising, and capital expenditure, and positively correlated with the likelihood of meeting or narrowly beating earnings targets. Johnson, Ryan, and Tian (2009) show that unrestricted stock is positively correlated with corporate fraud.

One potential solution to the potential negative consequences of short horizons is to extend the vesting period of equity. While the current debates surround the level of pay and the sensitivity of pay to performance, extending the horizon of incentives may be particularly valuable in overcoming moral hazard. However, lengthening vesting periods is not costless. First, doing so will potentially expose the CEO to risk outside his control. Second, Laux (2012) shows theoretically that, if the CEO forfeits unvested equity upon dismissal, he may engage in myopic actions to avoid the risk of dismissal until his equity has vested. Third, Brisley (2006) demonstrates that unvested equity ties up a significant portion of the CEO's wealth within the firm, and thus may cause him to turn down risky, value-creating projects.

We note that most of the potential remedies – indexation (where valuable), a crackdown on perks, updating contracts, and lengthening vesting periods – can be implemented by shareholders (or shareholder-aligned boards) themselves, without the need for regulatory intervention. The issue with regulation is that it is one-size-fits-all and cannot be adapted to a firm's particular circumstances. For example, a minimum vesting horizon of (say) 5 years for equity may be too short to induce investment in growth industries, and too long (thus subjecting the CEO to excessive risk) in mature industries. Indeed, Gopalan, Milbourn, Song and Thakor (2014) find that equity "duration" is longer in firms with higher growth opportunities, long-term assets, and R&D intensity, suggesting that the optimal vesting period varies according to firm characteristics. The recent increases in disclosure requirements, and say-on-pay legislation, are steps in the direction of allowing shareholders to ensure the optimality of contracts, as both give the information and ability to decide whether a given pay package is appropriate in a particular context. Moreover, any policy to reform executive pay should not focus narrowly on compensation alone, but recognize the systemic nature of the issue. Bolton, Scheinkman, and Xiong (2006) show that giving power to short-term shareholders may lead to them deliberately inducing myopia via short-term contracts. Thus, passing say-on-pay legislation alone may have little effect if shareholders do not have the incentives to set pay optimally. Encouraging shareholders to take long-term stakes (potentially through capital gains tax policy) will align their interests with long-run firm value. In addition, encouraging large holdings will ensure they have sufficient incentives to gather information on the firm's long-run value, rather than relying on public information such as short-term profit.

Regulatory intervention is, however, valuable if externalities exist. In Bénabou and Tirole (2015), competition causes firms to offer high incentives to screen for high-ability managers. However, strong incentives also lead to managers focusing excessively on measurable tasks and shirking on unmeasurable tasks, echoing Holmstrom and Milgrom (1991). Acharya and Volpin (2010) and Dicks (2012) show that competition can lead to spillover effects: if one firm overpays its workers (e.g. due to poor corporate governance), this will lead to other firms optimally doing so to remain competitive, even if they are well-governed.

5.2. Underexplored Areas

We start with potential avenues for future empirical analysis, before turning to ideas for theoretical research. There have been several high-impact empirical studies of executive compensation. Since the debate about the efficiency of pay concerns magnitudes, this is a field in which descriptive statistics alone are illuminating, for example Jensen and Murphy's (1990) and Hall and Liebman's (1998) seminal work on quantifying CEO incentives. Other studies have correlated CEO pay with outcomes such as firm value (e.g. Morck, Shleifer, and Vishny (1988), McConnell and Servaes (1990), and Himmelberg, Hubbard, and Palia (1999)). However, assigning causality is very difficult, as there are very few instruments for CEO incentives. Even the very basic question of whether CEO incentives positively affect firm value has not yet been satisfactorily answered. Thus, a first open question is to find good instruments for or quasi-exogenous shocks to CEO pay, to allow identification of the effects of incentives. There have been a limited number of attempts in this direction. Palia (2001) argues that CEO experience, education, and age, and firm volatility are instruments for executive compensation. Core and Larcker (2002) study increases in stock ownership mandated by CEOs approaching the minimum levels set by pre-announced guidelines. Shue and Townsend (2013) exploit the fact that options are granted according to multi-year cycles as an instrument for option grants. Edmans, Fang, and Lewellen (2015) and Edmans, Goncalves-Pinto, Wang, and Xu (2015) analyze the scheduled vesting of equity resulting from grants made several years prior.

Second, most empirical studies have been focused on public firms in the U.S., given the availability of the ExecuComp dataset. Research on compensation practices in private firms would be particularly useful. Since private firms are likely closer to the optimal contracting benchmark, due to the presence of a concentrated shareholder, such data would allow a comparison with similar public firms to assess whether pay in public firms represents rent extraction. For example, Cronqvist and Fahlenbrach (2013) study firms transitioning from public to private ownership. Additional studies investigating private firms in general (in addition to those that were formerly public) would be helpful. Another fruitful direction would be to study international data and analyze the determinants of cross-country differences in CEO pay. Conyon, Core, and Guay (2011) and Fernandes, Ferreira, Matos, and Murphy (2013) are useful steps in this direction. Moreover, while data on CEO wealth (an important determinant of both risk aversion and the private benefits from shirking) is typically unavailable in the U.S., it is sometimes available in other countries (see, e.g., Becker (2006)).

Third, structurally estimating a dynamic moral hazard model may allow us to study questions that are difficult to answer with reduced-form approaches. For example, it may allow us to quantify several important determinants of the optimal contract that are otherwise difficult to measure empirically, such as the CEO's risk aversion, cost of effort, ability to engage in manipulation, and desire for consumption smoothing. Relatedly, it can permit counterfactual analyses such as the effect on firm value of changes in these parameters, or how the possibility of myopia or short-termism changes the contract. In addition, formal joint tests of a theory's quantitative predictions can highlight where the theory fails, thus opening doors to future research. As examples of structural approaches, Gayle and Miller (2009) study the extent to which moral hazard can explain the rise in CEO pay. Margiotta and Miller (2000), who find that firms would suffer large losses from not contracting optimally and also estimate the gains that would arise in a first-best world where effort were observable. In addition, managers only require a small risk premium for the risk imposed by incentives – the benefits of incentives substantially outweigh the costs. Gayle and Miller (2015) show that moral hazard models in which managers can manipulate accounting reports better explain observed contracts than ones in which they cannot, and Gayle, Golan, and Miller (2015) decompose the sources of pay differences between large and small firms. Page (2011) quantitatively estimates the effect of increasing CEO ownership on firm value. Using a learning model, Taylor (2010) estimates the cost of firing that would rationalize observed turnover rates; a similar approach may uncover whether firing rates are optimal from a moral hazard perspective. Similarly, while existing tests of the rent extraction vs. optimal contracting hypotheses typically study the cross-section, an analysis of time-series dynamics would allow us to study whether the evolution of pay over time is consistent with optimal contracting.

We now move to open theoretical questions. First, most current market equilibrium models are static. It would be useful to add a dynamic moral hazard problem where incentives can be provided not only through contracts, but also by the threat of firing.[29] This will also allow us to understand what causes CEOs to move between firms. Moreover, the possibility of turnover adds complications to a standard dynamic model of moral hazard. The classic models of Lambert (1983) and Rogerson (1985) predict that the reduction in CEO pay caused by poor performance should be spread out over all future periods, to optimize risk sharing. However, the CEO may quit if future expected pay is low, reducing consumption smoothing possibilities.[30]

Second, while the "rent extraction" view has been influential, these arguments have been mainly stated verbally, e.g. Bebchuk and Fried (2004). It would be particularly useful to model the rent extraction view and compare its predictions to the data. One example is Kuhnen and Zwiebel (2009), where the manager has freedom to extract perks, but doing so reduces profits and thus shareholders' assessment of the manager's ability, which may lead to him being fired. The model predicts that perk consumption is increasing in production uncertainty (since it is easier to disguise low profits as resulting from a negative shock) and the manager's outside option (since firing is less of a concern). It is decreasing in uncertainty about the manager's ability, as then profits have a greater effect on shareholders' assessment of his ability and thus their firing decision. They find qualitative support for these predictions, measuring hidden pay with stock options, restricted stock, and annual pay not declared as salary and bonus. A further potential avenue in this line of research would be a rent extraction model that generates quantitative predictions, and allows for a horse-race between the two viewpoints.

Third, existing models of CEO pay are single-agent models, but CEOs work in teams where complementarities between agents exist. As a result, their contracts affect firm value not only directly through affecting the CEO's effort, but also indirectly because the CEO's effort level affects the optimal effort level set chosen by workers. This consideration in turn affects the optimal contract for the CEO. Separately, a team setting allows the study of the relative wages of the CEO and other employees, a question that has been of interest to regulators. Edmans, Goldstein, and Zhu (2013) analyze these issues within a CEO setting; Chen and Yoo (2001),

[29] See Eisfeldt and Kuhnen (2013) for a market equilibrium model with CEO turnover without moral hazard.

[30] The dynamic moral hazard models of DeMarzo and Sannikov (2006), DeMarzo and Fishman (2007), and Biais, Mariotti, Plantin, and Rochet (2007) assume risk-neutrality, and so consumption smoothing is a non-issue.

Kremer (1993), Winter (2004, 2006, 2010), and Gervais and Goldstein (2007) are analyses of contracting under production complementarities in general principal-agent settings.

Fourth, there has been substantial theoretical progress on continuous-time agency models which allow for the contracting problem to be solved with few assumptions. However, the empirical predictions of such models are typically less clear, given the absence of analytical solutions, and because numerical solutions depend on the parameters chosen. Future research may be able to identify clearer implications of these models, in particular comparative statics on how incentives and turnover-performance sensitivity should differ across firms.

Fifth, with few exceptions, existing executive compensation models are rational. Incorporating behavioral considerations has been successful in other fields of corporate finance (see the survey of Baker and Wurgler (2012)) and could be similarly fruitful here. Baker and Wurgler (2012) divide the behavioral corporate finance literature into two fields – managers who are irrational or have non-standard utility functions, and rational managers exploiting inefficient markets. As an example of the former, Dittmann, Maug, and Spalt (2010) show that incorporating loss aversion can explain the observed mix of stock and options, which standard utility functions cannot. As an example of the latter, Bolton, Scheinkman, and Xiong (2006) show that contracts that emphasize short-term performance may be a rational response to speculative markets. Other potential behavioral phenomena that could be incorporated into compensation models include bounded rationality, overconfidence (overweighting private signals and underweighting public signals), and optimism (overestimating one's own managerial ability or firm quality).

Finally, turning to a question for both theoretical and empirical research, we now have quantitative theories for the level of pay and "demand" side, given the supply of talent. However, we know relatively little on the "supply" side. Given the substantial pay premium that top executives command over other skilled professions (e.g. medicine or law), it would be interesting to study empirically the extent to which this premium results from limited supply, and if so, explore theoretically why supply appears to remain so limited – why more people do not enter the business profession. A related topic is to understand better the nature of the scarcity of CEO talent, e.g. whether it stems from innate skills, experience, lack of succession planning, and so on.[31]

6. Conclusion

This article has presented a number of optimal contracting models of executive compensation under a unifying framework. We commenced with assignment models of the CEO labor market. More talented managers are matched with larger firms, since their talent is scalable. Their talent also allows them to command higher wages, leading to quantitative predictions for the cross-

[31] For the supply of skills of general workers, see Goldin and Katz (2011).

sectional relationship between pay and firm size. Since the dollar benefits of talent are greater in larger firms, the model also implies that pay should rise over time as average firm size grows.

We then moved to static moral hazard models, and showed that the correct empirical measure of incentives depends on whether we believe effort has additive or multiplicative effects on firm value and CEO utility. Moreover, if the effect of effort scales with firm size, dollar-dollar incentives should optimally be weaker in smaller firms. If effort is continuous and the optimal effort level is interior and endogenous, as in Holmstrom and Milgrom (1987), incentives should be increasing in the benefits of effort and decreasing in risk, risk aversion, and the cost of effort. In contrast, if effort is binary, or continuous and the principal wishes to implement full productive efficiency, risk and risk aversion do not affect incentives but increase the level of pay. The rarity of relative performance evaluation appears to contradict the Holmstrom (1979) informativeness principle, but we discussed potential rationalizations of this practice. If the principal aims to induce risk-taking as well as effort from the CEO, the contract should be convex and generally contain options, in contrast to the linear contracts advocated by Holmstrom and Milgrom (1987).

We finally discussed dynamic moral hazard models. To achieve optimal risk-sharing, the reward for good performance should be smoothed over future periods. As the CEO approaches retirement, since there are fewer periods over which to engage in smoothing, the sensitivity of current pay to performance should rise. If the CEO can engage in private saving, his wage profile is typically backward-loaded, to remove such saving incentives.

While each model has different features and tackles different questions, we highlight two common themes. First, the modeling assumptions (e.g. whether preferences or production functions are additive or multiplicative, whether actions are continuous or discrete, and whether private saving is possible) can have significant impact on the model's predictions. Second, we emphasize the empirical predictions of the models and compared them to the data. In particular, observed practices that are often interpreted as evidence of rent extraction are also qualitatively, and sometimes quantitatively, consistent with optimal contracting. However, consistency with observed correlations does not prove that real-life practices are optimal; instead, it merely emphasizes caution before attempting to intervene by regulation. Whether observed contracts result from efficiency or rent extraction is still an open question, and we highlight other potential avenues for future research. We look forward to seeing this literature continue to evolve.

References

Acemoglu, Daron, and David Autor. 2012. "What Does Human Capital Do? A Review of Goldin and Katz's The Race between Education and Technology." *Journal of Economic Literature* 50 (2): 426–463.

Acharya, Viral V., and Paolo F. Volpin. 2010. "Corporate Governance Externalities." *Review of Finance* 14 (1): 1–33.

Acharya, Viral V., Kose John, and Rangarajan K. Sundaram. 2000. "On the Optimality of Resetting Executive Stock Options." *Journal of Financial Economics* 57 (1): 65–101.

Aggarwal, Rajesh K. and Andrew A. Samwick. 1999. "The Other Side of the Trade-off: The Impact of Risk on Executive Compensation." *Journal of Political Economy* 107 (1): 65–105.

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

Anantharaman, Divya, Vivian W. Fang, and Guojin Gong. 2014. "Inside Debt and the Design of Corporate Debt Contracts." *Management Science* 60 (5): 1260–1280.

Arnott, Richard, and Joseph E. Stiglitz. 1988. "Randomization with Asymmetric Information." *RAND Journal of Economics* 19 (3): 344–362.

Axtell, Robert L. 2001. "Zipf Distribution of US Firm Sizes." *Science* 293 (5536): 1818–1820.

Baker, George P., Michael C. Jensen and Kevin J. Murphy. 1988. "Compensation and Incentives: Practice vs. Theory." *Journal of Finance* 43 (3): 593–616.

Baker, George P. 1992. "Incentive Contracts and Performance Measurement." *Journal of Political Economy* 100 (3): 598–614.

Baker, Malcolm and Jeffrey Wurgler, "Behavioral Corporate Finance: An Updated Survey." *Handbook of the Economics of Finance*, Vol. 2. New York and Oxford: Elsevier/North-Holland: 357–424.

Bakija, Jon, Adam Cole, and Bradley T. Heim (2012): "Jobs and Income Growth of Top Earners and the Causes of Changing Income Inequality: Evidence From U.S. Tax Return Data." Williams College Working Paper.

Baranchuk, Nina, Glenn MacDonald, and Jun Yang (2011): "The Economics of Super Managers." *Review of Financial Studies* 24 (10): 3321–3368.

Bebchuk, Lucian A., K. J. Martijn Cremers, and Urs Peyer. 2011. "The CEO Pay Slice." *Journal of Financial Economics* 102 (1): 199–221.

Bebchuk, Lucian A., and Jesse M. Fried. 2004. *Pay Without Performance: The Unfulfilled Promise of Executive Compensation*. Cambridge, MA: Harvard University Press.

Becker, Bo. 2006. "Wealth and Executive Compensation." *Journal of Finance* 61 (1): 379–397.

Benmelech, Efraim, Eugene Kandel, and Pietro Veronesi. 2010. "Stock-Based Compensation and CEO (Dis)Incentives." *The Quarterly Journal of Economics* 125 (4): 1769–1820.

Bénabou, Roland and Jean Tirole. 2015. "Bonus Culture: Competitive Pay, Screening, and Multitasking." *Journal of Political Economy*, forthcoming.

Bennedsen, Morten, Francisco Perez-Gonzalez, and Daniel Wolfenzon. 2010. "Do CEOs Matter?" Copenhagen Business School Working Paper.

Bereskin, Frederick and David C. Cicero. 2013. "CEO Compensation Contagion: Evidence from an Exogenous Shock." *Journal of Financial Economics* 107 (2): 477–493.

Bertrand, Marianne, and Sendhil Mullainathan. 2001. "Are CEOs Rewarded for Luck? The Ones Without Principals Are." *Quarterly Journal of Economics* 116 (3): 901–932.

Biais, Bruno, Thomas Mariotti, Jean-Charles Rochet, and Stephane Villeneuve. 2010. "Large Risks, Limited Liability, and Dynamic Moral Hazard." *Econometrica* 78 (1): 73–118.

Biais, Bruno, Thomas Mariotti, Guillaume Plantin, and Jean-Charles Rochet. 2007. "Dynamic Security Design: Convergence to Continuous Time and Asset Pricing Implications." *Review of Economic Studies* 74 (2): 345–390.

Bolton, Patrick, and Mathias Dewatripont. 2005. *Contract Theory* (MIT Press, Cambridge, MA).

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

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

Brander, James A., and Michel Poitevin. 1992. "Managerial Compensation and the Agency Costs of Debt Finance." *Managerial and Decision Economics* 13 (1): 55–64.

Brenner, Menachem, Rangarajan K. Sundaram, and David Yermack. 2000. "Altering the Terms of Executive Stock Options." *Journal of Financial Economics* 57 (1): 103–28.

Brisley, Neil. 2006. "Executive Stock Options: Early Exercise Provisions and Risk-Taking Incentives." *Journal of Finance* 61 (5): 2487–2509.

Bushman, Robert M., Raffi J. Indjejikian and Abbie Smith. 1996. "CEO Compensation: The Role of Individual Performance Evaluation." *Journal of Accounting and Economics* 21 (2): 161–193.

Carpenter, Jennifer N. 2000. "Does Option Compensation Increase Managerial Risk Appetite?" *Journal of Finance* 55 (5): 2311–2332.

Carpenter, Jennifer, Richard Stanton and Nancy Wallace. 2013. "The Importance of Behavioral Factors in the Exercise and Valuation of Employee Stock Options,", Working Paper.

Carroll, Gabriel. 2012. "When Are Local Incentive Constraints Sufficient?" *Econometrica* 80 (2): 661–686.

Cassell, Cory A., Shawn X. Huang, Juan Manuel Sanchez, and Michael D. Stuart. 2012. "Seeking Safety: The Relation Between CEO Inside Debt Holdings and the Riskiness of Firm Investment and Financial Policies." *Journal of Financial Economics* 103 (3): 588–610.

Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb. 2015a. "The Value of Informativenss for Contracting." HEC Montreal Working Paper

Chaigneau, Pierre, Alex Edmans, and Daniel Gottlieb. 2015b. "The Informativeness Principle Under Limited Liability." HEC Montreal Working Paper

Che, Yeon-Koo, and Seung-Weon Yoo. 2001. "Optimal Incentives for Teams." *American Economic Review* 91(3): 525–541.

Cheng, Ing-Haw, Harrison Hong and José A. Scheinkman. 2015. "Yesterday's Heroes: Compensation and Risk at Financial Firms." *Journal of Finance*, forthcoming.

Coles, Jeffrey L., Naveen D. Daniel and Lalitha Naveen. 2006. "Managerial Incentives and Risk-Taking." *Journal of Financial Economics* 79 (2): 431–468.

Conyon, Martin J., and Kevin J. Murphy. 2000. "The Prince and the Pauper? CEO Pay in the United States and United Kingdom." *Economic Journal* 110 (467): F640–F671.

Conyon, Martin J., John E. Core and Wayne R, Guay. 2011. "Are U.S. CEOs Paid More than UK CEOs? Inferences from Risk-Adjusted Pay." *Review of Financial Studies* 24 (2): 402–38.

Cooley, Thomas F., and Edward C. Prescott. 1995. "Economic Growth and Business Cycles" in Thomas Cooley (ed.), *Frontiers in Business Cycle Research*. Princeton, NJ: Princeton University Press.

Core, John E., and Wayne R. Guay. 1999. "The Use of Equity Grants to Manage Optimal Equity Incentive Levels." *Journal of Accounting and Economics* 28 (2): 151–184.

Core, John E., and David F. Larcker. 2002. "Performance Consequences of Mandatory Increases in Executive Stock Ownership." *Journal of Financial Economics* 64 (3): 317–340.

Core, John E., Wayne R. Guay, and David F. Larcker. 2003. "Executive Equity Compensation and Incentives: A Survey." *Federal Reserve Bank of New York Economic Policy Review* 9 (1): 27–50.

Cornelli, Francesca, Zbigniew Kominek, and Alexander Ljungqvist. 2013. "Monitoring Managers: Does It Matter?" *Journal of Finance* 68 (2): 431–481.

Coughlan, Anne T. and Ronald M. Schmidt. 1985. "Executive Compensation, Management Turnover and Firm Performance: An Empirical Investigation." *Journal of Accounting and Economics* 7 (1): 43–66.

Cremers, K. J. Martijn, and Yaniv Grinstein. 2015. "Does the Market for CEO Talent Explain Controversial CEO Pay Practices?" *Review of Finance*, forthcoming.

Cronqvist, Henrik, and Rüdiger Fahlenbrach. 2013. "CEO Contract Design: How Do Strong Principals Do It?" *Journal of Financial Economics* 108 (3): 659–74.

DeMarzo, Peter M., and Michael J. Fishman. 2007. "Agency and Optimal Investment Dynamics." *Review of Financial Studies* 20 (1): 151–188.

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

Demsetz, Harold, and Kenneth Lehn. 1985. "The Structure of Corporate Ownership:

Causes and Consequences." *Journal of Political Economy* 93 (6): 1155–1177.

Dicks, David. 2012. "Executive Compensation and the Role for Corporate Governance Regulation." *Review of Financial Studies* 25 (6): 1971–2004.

Dittmann, Ingolf, and Ernst Maug. 2007. "Lower Salaries and No Options? On the Optimal Structure of Executive Pay." *Journal of Finance* 62 (1): 303–343.

Dittmann, Ingolf, Ernst Maug, and Oliver Spalt. 2010. "Sticks or Carrots? Optimal CEO Compensation When Managers Are Loss-Averse." *Journal of Finance* 65 (6): 2015–2050.

Dittmann, Ingolf, Ernst Maug and Oliver Spalt. 2013. "Indexing Executive Compensation Contracts." *Review of Financial Studies* 26 (12), 3182–3224.

Dittmann, Ingolf, and Ko-Chia Yu. 2011. "How Important Are Risk-Taking Incentives in Executive Compensation?" Erasmus University Rotterdam Working Paper.

Edmans, Alex, Vivian W. Fang, and Katharina A. Lewellen. 2015. "Equity Vesting and Managerial Myopia." London Business School Working Paper.

Edmans, Alex, and Xavier Gabaix. 2009. "Is CEO Pay Really Inefficient? A Survey of New Optimal Contracting Theories." *European Financial Management* 15 (3), 486–496.

Edmans, Alex, and Xavier Gabaix. 2011a. "The Effect of Risk on the CEO Market." *Review of Financial Studies* 24 (8): 2822–2863.

Edmans, Alex, and Xavier Gabaix. 2011b. "Tractability in Incentive Contracting." *Review of Financial Studies* 24 (9): 2865–2894.

Edmans, Alex, Xavier Gabaix, and Augustin Landier. 2009. "A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium." *Review of Financial Studies* 22 (12):4881–4917.

Edmans, Alex, Xavier Gabaix, Tomasz Sadzik, and Yuliy Sannikov. 2012. "Dynamic CEO Compensation." *Journal of Finance* 67 (5), 1603–1647.

Edmans, Alex, Itay Goldstein, and John Y. Zhu. 2013. "Contracting With Synergies." London Business School Working Paper.

Edmans, Alex, Luis Goncalves-Pinto, Yanbo Wang, and Moqi Xu. 2015. "Strategic News Releases in Equity Vesting Months." London Business School Working Paper.

Edmans, Alex and Qi Liu. 2010. "Inside Debt." *Review of Finance* 15 (1): 75–102.

Eisfeldt, Andrea, and Camelia M. Kuhnen. 2013. "CEO Turnover in a Competitive Assignment Framework." *Journal of Financial Economics* 109 (2): 351–372.

Farhi, Emmanuel, and Ivan Werning. 2012. "Capital Taxation: Quantitative Explorations of the Inverse Euler Equation." *Journal of Political Economy* 102 (3), 398–445.

Fernandes, Nuno, Miguel A. Ferreira, Pedro Matos, and Kevin J. Murphy. 2013. "Are U.S. CEOs Paid More? New International Evidence." *Review of Financial Studies* 26 (2): 323–67.

Frydman, Carola, and Raven Saks. 2010. "Executive Compensation: A New View From a Long-Term Perspective, 1936-2005." *Review of Financial Studies* 23, 2099–2138.

Frydman, Carola, and Dirk Jenter. 2010. "CEO Compensation." *Annual Review of Financial Economics* 2: 75–102.

Frydman, Carola. 2007. "Rising Through the Ranks: The Evolution of the Market for Corporate Executives, 1936-2003." Boston University Working Paper.

Gabaix, Xavier, and Augustin Landier. 2008. "Why Has CEO Pay Increased So Much?" *Quarterly Journal of Economics* 123 (1): 49–100.

Gabaix, Xavier, Augustin Landier, and Julien Sauvagnat. 2014. "CEO Pay and Firm Size: An Update After the Crisis." *The Economic Journal* 124 (574): F40–F59

Garen, John E. 1994. "Executive Compensation and Principal-Agent Theory." *Journal of Political Economy* 102 (6): 1175–1199.

Garicano, Luis and Esteban Rossi-Hansberg. 2006. "Organization and Inequality in a Knowledge Economy." *Quarterly Journal of Economics* 121 (4): 1383–1435.

Garrett, Daniel, and Alessandro Pavan. 2012. "Managerial Turnover in a Changing World." *Journal of Political Economy* 102 (5): 879–925.

Garrett, Daniel, and Alessandro Pavan. 2015. "Dynamic Managerial Compensation: On the Optimality of Seniority-Based Schemes." *Journal of Economic Theory*, forthcoming.

Gayle, George-Levi and Robert A. Miller. 2009. "Has Moral Hazard Become a More Important Factor in Managerial Compensation?" *American Economic Review* 99 (5): 1740–69.

Gayle, George-Levi and Robert A. Miller. 2015. "Identifying and Testing Models of Managerial Compensation." *Review of Economic Studies*, forthcoming.

Gayle, George-Levi, Limor Golam, and Robert A. Miller. 2015. "Promotion, Turnover, and Compensation in the Executive Labor Market." *Econometrica*, forthcoming.

Gervais, Simon, and Itay Goldstein. 2007. "The Positive Effects of Biased Self-Perceptions in Firms." *Review of Finance* 11 (3): 453–496.

Gibbons, Robert, and Kevin J. Murphy. 1992. "Optimal Incentive Contracts in the Presence of Career Concerns: Theory and Evidence." *Journal of Political Economy* 100 (3): 468–505.

Gjesdal, Frøystein. 1982. "Information and Incentives: The Agency Information Problem." *Review of Economic Studies* 49 (3): 373–390.

Goldin, Claudia Dale, and Lawrence F. Katz. 2009. *The Race Between Education and Technology*. Cambridge, MA: Harvard University Press

Goldman, Eitan, and Steve Slezak. 2006. "An Equilibrium Model of Incentive Contracts in the Presence of Information Manipulation." *Journal of Financial Economics* 80 (3): 603–26.

Gong, Guojin, Laura Yue Li, and Jae Yong Shin. 2011. "Relative Performance Evaluation and Related Peer Groups in Executive Compensation Contracts." *The Accounting Review* 86 (3): 1007–1043.

Gopalan, Radhakrishnan, Todd Milbourn and Fenghua Song. 2010. "Strategic Flexibility and the Optimality of Pay for Sector Performance." *Review of Financial Studies* 23(5): 2060–98.

Gopalan, Radhakrishnan, Todd Milbourn, Fenghua Song, and Anjan V. Thakor. 2014. "Duration of Executive Compensation." *Journal of Finance* 69 (6): 2777–2817.

Grossman, Sanford J., and Oliver D. Hart. 1983. "An Analysis of the Principal-Agent Problem." *Econometrica* 51 (1): 7–45

Hall, Brian J., and Jeffrey B. Liebman. 1998. "Are CEOs Really Paid Like Bureaucrats?" *Quarterly Journal of Economics* 112 (3): 653–691.

Hall, Brian J., and Kevin J. Murphy. 2000. "Optimal Exercise Prices for Executive Stock Options." *American Economic Review* 90 (2): 209–214.

Hall, Brian J., and Kevin J. Murphy. 2002. "Stock Options for Undiversified Executives." *Journal of Accounting and Economics* 33 (1): 3–42.

Hall, Brian J., and Thomas A. Knox. 2004. "Underwater Options and the Dynamics of Executive Pay-to-Performance Sensitivities." *Journal of Accounting Research* 42 (2): 365–412.

Harris, Milton, and Bengt Holmstrom. 1982. "A Theory of Wage Dynamics." *Review of Economic Studies* 49 (3): 315–333.

Harris, Milton, and Artur Raviv. 1979. "Optimal Incentive Contracts with Imperfect Information." *Journal of Economic Theory* 20 (2): 231–259.

He, Zhiguo. 2012. "Dynamic Compensation Contracts with Private Savings." *Review of Financial Studies* 25 (5): 1494–1549.

Hellwig, Martin, and Klaus Schmidt. 2002. "Discrete-Time Approximations of the Holmstrom-Milgrom Brownian-Motion Model of Intertemporal Incentive Provision." *Econometrica* 70 (6): 2225–2264.

Hermalin, Benjamin E. 2005. "Trends in Corporate Governance." *Journal of Finance* 60 (5): 2351–2384.

Hermalin, Benjamin E., and Michael S. Weisbach. 1998. "Endogenously Chosen Boards of Directors and Their Monitoring of the CEO." *American Economic Review* 88 (1): 96–118.

Hermalin, Benjamin E., and Michael S. Weisbach. 2012. "Information Disclosure and Corporate Governance." *Journal of Finance* 67 (1): 195–233.

Himmelberg, Charles, R. Glenn Hubbard, and Darius Palia. 1999. "Understanding the Determinants of Managerial Ownership and the Link Between Ownership and Performance." *Journal of Financial Economics* 53 (3): 353–84.

Holmstrom, Bengt. 1979. "Moral Hazard and Observability." *Bell Journal of Economics* 10 (1): 74–91.

Holmstrom, Bengt. 1999. "Managerial Incentive Problems: A Dynamic Perspective." *Review of Economic Studies* 66 (1): 169–181.

Holmstrom, Bengt, and Paul Milgrom. 1987. "Aggregation and Linearity in the Provision of Intertemporal Incentives." *Econometrica* 55 (2): 308–28.

Holmstrom, Bengt, and Jean Tirole. 1997. "Financial Intermediation, Loanable Funds, and the Real Sector." *Quarterly Journal of Economics* 112 (3): 663–691.

Inderst, Roman, and Holger M. Mueller. 2010. "CEO Replacement Under Private Information." *Review of Financial Studies* 23 (8): 2935-2969.

Innes, Robert D. 1990. "Limited Liability and Incentive Contracting with Ex-Ante Action Choices." *Journal of Economic Theory* 52 (1): 45–67.

Ittner, Christopher D., David F. Larcker, and Madhav V. Rajan. 1997. "The Choice of Performance Measures in Annual Bonus Contracts." *The Accounting Review* 72 (2): 231–255.

Jensen, Michael C., and William Meckling. 1976. "Theory of the Firm: Managerial Behavior, Agency Costs, and Capital Structure." *Journal of Financial Economics* 3 (4): 305–360.

Jensen, Michael C., and Kevin J. Murphy. 1990. "Performance Pay and Top-Management Incentives." *Journal of Political Economy* 98 (2): 225–64.

Jensen, Michael C., and Kevin J. Murphy. 2004. "Remuneration: Where We've Been, How We Got to Here, What are the Problems, and How to Fix Them." Harvard University Working Paper.

Jenter, Dirk and Fadi Kanaan. 2015. "CEO Turnover and Relative Performance Evaluation." *Journal of Finance*, forthcoming.

Jenter, Dirk and Katharina A. Lewellen. 2014. "Performance-Induced CEO Turnover." Stanford University Working Paper.

Jewitt, Ian. 1988. "Justifying the First-Order Approach to Principal-Agent Problems." *Econometrica* 56 (5): 1177–1190.

Jin, Li. 2002. "CEO Compensation, Diversification, and Incentives." *Journal of Financial Economics* 66 (1): 29–63.

Johnson, Shane, Harley E. Ryan and Yisong Tian. 2009. "Managerial Incentives and Corporate Fraud: The Sources of Incentives Matter." *Review of Finance* 13 (1): 115–145.

Joskow, Paul, John R. Meyer, Nancy Rose, Andrea Shepard, and Sam Peltzman. 1993. "Regulatory Constraints on CEO Compensation." *Brookings Papers on Economic Activity. Microeconomics* 1993 (1): 1–72.

Jovanovic, Boyan. 1979. "Job Matching and the Theory of Turnover." *Journal of Political Economy* 87 (5): 972–90.

Kaplan, Steven N. 2012. "Executive Compensation and Corporate Governance in the U.S.: Perceptions, Facts, and Challenges." *Cato Papers on Public Policy* 2, 99–157.

Kaplan, Steven N., and Bernadette Minton. 2011. "How Has CEO Turnover Changed?" *International Review of Finance* 12 (1): 57-87.

Kaplan, Steven N., and Joshua Rauh. 2010. "Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?" *Review of Financial Studies* 23 (3): 1004–1050.

Kostuik, Peter F. 1990. "Firm Size and Executive Compensation." *Journal of Human Resources* 25 (1): 90–105.

Kremer, Michael. 1993. "The O-Ring Theory of Economic Development." *Quarterly Journal of Economics* 108 (3): 551–575.

Kuhnen, Camelia M. and Jeffrey Zwiebel. 2009. "Executive Pay, Hidden Compensation, and Managerial Entrenchment." University of North Carolina Working Paper.

Lacker, Jeffrey, and John Weinberg. 1989. "Optimal Contracts Under Costly State Falsification." *Journal of Political Economy* 97 (6): 1345–1363.

Lambert, Richard. 1983. "Long-Term Contracts and Moral Hazard." *Bell Journal of Economics* 14 (2): 441–452.

Lambert, Richard A., and David F. Larcker. 1987. "An Analysis of the Use of Accounting and Market Measures of Performance in Executive Compensation Contracts." *Journal of Accounting Research* 25: 85–125.

Lambert, Richard A., David F. Larcker, and Robert Verrecchia. 1991. "Portfolio Considerations in Valuing Executive Compensation." *Journal of Accounting Research* 29 (1): 129–149.

Laux, Volker. 2012. "Stock Option Vesting Conditions, CEO Turnover, and Myopic Investment." *Journal of Financial Economics* 106 (3): 513–526.

Lazear, Edward P. 1979. "Why Is There Mandatory Retirement?" *Journal of Political Economy* 87 (6): 1261–1284.

Lie, Erik. 2005. "On the Timing of CEO Stock Option Awards." *Management Science* 51 (5): 802–812.

Lucas, Robert E., Jr. 1978. "On the Size Distribution of Business Firms." *Bell Journal of Economics* 9 (2): 508–523.

Manso, Gustavo. 2011. "Motivating Innovation." *Journal of Finance* 66 (5): 1823–1860.

Margiotta, Mary M., and Robert A. Miller. 2000. "Managerial Compensation and the Cost of Moral Hazard." *International Economic Review* 41 (3): 669–719.

McConnell, John J., and Henri Servaes. 1990. "Additional Evidence on Equity Ownership and Corporate Value." *Journal of Financial Economics* 27 (2): 595–612.

Morck, Randall, Andrei Shleifer and Robert W. Vishny. 1988. "Management Ownership and Market Valuation: An Empirical Analysis." *Journal of Financial Economics* 20 (1-2): 293–315.

Murphy, Kevin J. 1985. "Corporate Performance and Managerial Remuneration: An Empirical Analysis." *Journal of Accounting and Economics* 7 (1): 11–42.

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

Murphy, Kevin J. 2013. "Executive Compensation: Where We Are, and How We Got There" in George Constantinides, Milton Harris, and René Stulz (eds.). *Handbook of the Economics of Finance*. Elsevier/North-Holland.

Murphy, Kevin J., and Jan Zabojnik. 2006. "Managerial Capital and the Market for CEOs." Working Paper, University of Southern California.

Nagel, Gregory L. 2010. "The Effect of Labor Market Demand on U.S. CEO Pay Since 1980." *Financial Review* 45 (4): 931–950.

Oyer, Paul. 2004. "Why Do Firms Use Incentives That Have No Incentive Effects?" *Journal of Finance* 59 (4): 1619–1650.

Oyer, Paul, and Scott Schaefer. 2005. "Why Do Some Firms Give Stock Options to All Employees?: An Empirical Examination of Alternative Theories." *Journal of Financial Economics*

76 (1): 99–133.

Page, Beau. 2011. "CEO Ownership and Firm Value: Evidence From a Structural Estimation." Working Paper, University of Houston.

Palia, Darius. 2001. "The Endogeneity of Managerial Compensation in Firm Valuation: A Solution." *Review of Financial Studies* 14 (3): 735–764.

Peng, Lin, and Ailsa Roell. 2008. "Manipulation and Equity-Based Compensation." *American Economic Review* 98 (2): 285–290.

Peng, Lin, and Ailsa Roell. 2014. "Managerial Incentives and Stock Price Manipulation." *Journal of Finance* 69 (2): 487–526.

Peters, Florian S., and Alexander F. Wagner. 2014. "The Executive Turnover Risk Premium." *Journal of Finance* 69 (4): 1529–1563.

Piketty, Thomas. 2014. *Capital in the 21st century*, Harvard University Press.

Piketty, Thomas, and Emmanuel Saez. 2003. "Income Inequality in the United States, 1913-1998." *Quarterly Journal of Economics* 118 (1): 1-39

Prendergast, Canice. 1999. "The Provision of Incentives in Firms." *Journal of Economic Literature* 37 (1): 7–63.

Prendergast, Canice. 2002. "The Tenuous Trade-off Between Risk and Incentives." *Journal of Political Economy* 110 (5): 1071–1102.

Rogerson, William P. 1985. "The First Order Approach to Principal-Agent Problems." *Econometrica* 53 (6): 1357–1368

Rose, Nancy, and Andrea Shepard. 1997. "Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?" *RAND Journal of Economics* 28 (3): 489–514.

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

Ross, Stephen A. 2004. "Compensation, Incentives, and the Duality of Risk Aversion and Riskiness." *Journal of Finance* 59 (1): 207–225.

Sannikov, Yuliy. 2008. "A Continuous-Time Version of the Principal-Agent Problem." *Review of Economic Studies* 75 (3): 957–984.

Sappington, David. 1983. "Limited Liability Contracts Between Principal and Agent." *Journal of Economic Theory* 29 (1): 1–21.

Sattinger, Michael. 1993. "Assignment Models of the Distribution of Earnings." *Journal of Economic Literature* 31 (2): 831–880.

Shue, Kelly, and Richard Townsend. 2013. "Swinging for the Fences: Executive Reactions to Quasi-Random Option Grants." University of Chicago Working Paper.

Shue, Kelly, and Richard Townsend. 2015. "Growth through Rigidity: An Explanation of the Rise in CEO Pay." University of Chicago Working Paper.

Smith, Clifford W., and Rene M. Stulz. 1985. "The Determinants of Firms' Hedging Policies." *Journal of Financial and Quantitative Analysis* 20 (4): 391-405.

Stein, Jeremy C. 1988. "Takeover Threats and Managerial Myopia." *Journal of Political Economy* 46 (1): 61–80.

Sundaram, Rangarajan and David L. Yermack. 2007. "Pay Me Later: Inside Debt and Its Role in Managerial Compensation." *Journal of Finance* 62 (4): 1551-1588.

Taylor, Lucian A. 2010. "Why Are CEOs Rarely Fired? Evidence from Structural Estimation." *The Journal of Finance* 65 (6): 2051–2087.

Taylor, Lucian A. 2013. "CEO Wage Dynamics: Estimates from a Learning Model." *Journal of Financial Economics* 108 (1): 79–98.

Terviö, Marko. 2008. "The Difference that CEOs Make: An Assignment Model Approach." *American Economic Review* 98 (3): 642–668.

Tirole, Jean. 2005. *The Theory of Corporate Finance.* Princeton, NJ: Princeton University Press.

Tirole, Jean. 2009. "Cognition and Incomplete Contracts." *American Economic Review* 99 (1): 265–294.

Warner, Jerold B., Ross L. Watts, and Karen H. Wruck. 1988. "Stock Prices and Top Management Changes." *Journal of Financial Economics* 20 (1-2): 461–492.

Wei, Chenyang, and David Yermack. 2011. "Investor Reactions to CEOs' Inside Debt Incentives." *Review of Financial Studies* 24 (11): 3813–3840.

Weisbach, Michael. 1988. "Outside Directors and CEO Turnover." *Journal of Financial Economics* 20 (1-2): 431–460.

Winter, Eyal. 2004. "Incentives and Discrimination." *American Economic Review* 94 (3): 764–773.

Winter, Eyal. 2006. "Optimal Incentives for Sequential Production Processes." *RAND Journal of Economics* 37 (2): 376–390.

Winter, Eyal. 2010. "Transparency and Incentives Among Peers." *RAND Journal of Economics* 41 (3): 504–523.

Yang, Jun. 2009. "Timing of Effort and Reward: Three-Sided Moral Hazard in a Continuous-Time Model." *Management Science* 56 (9): 1568–1583.

Yermack, David. 1995. "Do Corporations Award CEO Stock Options Effectively?" *Journal of Financial Economics* 39 (2): 237–269.

Yermack, David. 1997. "Good Timing: CEO Stock Option Awards and Company News Announcements." *Journal of Finance* 52 (2): 449–476.

Yermack, David. 2006. "Golden Handshakes: Separation Pay for Retired and Dmismissed CEOs." *Journal of Accounting and Economics* 41 (3): 237–256.

Zhu, John Y. 2014. "Myopic Agency." University of Pennsylvania Working Paper.

A. Longer Derivations

A.1. Proof of two core results on dynamic incentives

Heuristic proof of the Martingal representation theorem (69) We show two proofs – one rigorous, one heuristic that shows the economic origin of the result.

Heuristic proof. We reason by keeping the dt terms, dropping the $O\left(dt^2\right)$ terms.

$$\begin{aligned}
U_t &= E_t\left[\int_t^{\infty} e^{-\delta(s-t)} K_s ds\right] \\
&= E_t\left[\int_t^{t+dt} e^{-\delta(s-t)} K_s ds\right] + E_t\left[\int_{t+dt}^{\infty} e^{-\delta(s-t)} K_s ds\right] \\
&= K_t dt + o\left(dt\right) + e^{-\delta dt} E_t\left[E_{t+dt}\int_{t+dt}^{\infty} e^{-\delta(s-(t+dt))} K_s ds\right] \\
&= K_t dt + o\left(dt\right) + e^{-\delta dt} E_t U_{t+dt} \\
&= K_t dt + o\left(dt\right) + \left(1-\delta dt\right)\left(U_t + E_t dU_t\right) + o\left(dt\right) \\
&= U_t + \left(K_t - \delta U_t\right) dt + E_t\left[dU_t\right] + o\left(dt\right)
\end{aligned}$$

Hence,

$$0 = \left(K_t - \delta U_t\right) dt + E_t\left[dU_t\right] \tag{77}$$

As $dU_t - E_t\left[dU_t\right]$ has mean 0, and the only source of randomness is dZ_t,it makes sense that it can be written

$$dU_t - E_t\left[dU_t\right] = \xi_t dZ_t \tag{78}$$

for some adapted process ξ_t. Combining this with (77), we obtain (69).

Rigorous proof. The following proof (after Sannikov 2008) is more rigorous, but a bit less explicit about the origins of the result. Define

$$V_\infty := \int_0^\infty e^{-\delta s} K_s ds$$

and $V_t = E_t\left[V_\infty\right]$. Then, V_t is a martingale. That implies that we can write $dV_t = \xi_t e^{-\delta t} dZ_t$ for some adapted process ξ_t. Hence, $V_t = V_0 + \int_0^t \xi_s e^{-\delta s} dZ_s$. On the other hand,

$$\begin{aligned}
V_t = E_t\left[V_\infty\right] &= \int_0^t e^{-\delta s} K_s ds + E_t \int_t^\infty e^{-\delta s} K_s ds \\
&= \int_0^t e^{-\delta s} K_s ds + e^{-\delta t} U_t
\end{aligned}$$

as $U_t = E_t \int_t^\infty e^{-\delta(s-t)} K_s ds$. Differentiating w.r.t. t, we obtain:

$$dV_t = \xi_t e^{-\delta t} dZ_t = e^{-\delta t} K_t dt + e^{-\delta t} dU_t - \delta e^{-\delta t} U_t dt$$

hence

$$dU_t = (\delta U_t - K_t)\, dt + \xi_t dZ_t.$$

□

Heuristic proof of the Hamilton-Jacobi Bellman ("HJB") equation (70). The result is standard, but here we provide a heuristic proof. We first ignore the maximization over C. We have, as in (77),

$$0 = (f(x_t, C_t) - \delta Q_t)\, dc + E_t[dQ_t]$$

Now, by Ito's lemma using $Q_t = Q(x_t)$, $E_t[dQ_t]/dt = Q'(x_t)\mu(x_t, C) + \frac{1}{2}Q''(x_t)\sigma^2(x_t, C)$. Hence,

$$0 = f(x_t, C_t) - \delta Q(x_t) + Q'(x_t)\mu(x_t, C_t) + \frac{1}{2}Q''(x_t)\sigma^2(x_t, C_t)$$

Next, the principal will want to maximize over C the right-hand side (i.e. maximization $Q(x_t)$), hence (70). □

A.2. Proof of other results

Heuristic proof of (58)-(61) See Edmans, Gabaix, Sadzik, and Sannikov (2012) for a rigorous proof. We present a heuristic proof in a simple case that conveys the key intuition. We consider $L = T = 2$ and impose the PS constraint. We wish to show that the optimal contract is given by

$$\ln c_1 = g'(a^*)\frac{r_1}{2} + \kappa_1, \quad \ln c_2 = g'(a^*)\left(\frac{r_1}{2} + r_2\right) + \kappa_1 + k_2 \tag{79}$$

for some constants κ_1 and k_2 that make the participation constraint bind.

Step 1: Optimal log-linear contract

We first solve the problem in a restricted class where contracts are log-linear, that is,

$$\ln c_1 = \theta_1 r_1 + \kappa_1,\ \ln c_2 = \theta_{21} r_1 + \theta_2 r_2 + \kappa_1 + k_2 \tag{80}$$

for some constants θ_1, θ_{21}, θ_2, κ_1, k_2. This first step is not necessary but clarifies the economics, and in more complex cases is helpful to guess the form of the optimal contract.

First, intuitively, the optimal contract entails consumption smoothing, that is, shocks to consumption are permanent. This observation implies $\theta_{21} = \theta_1$. To prove this, the PS constraint (61) yields

$$1 = E_1\left[\frac{c_1}{c_2}\right] = e^{(\theta_1 - \theta_{21})r_1} E_1\left[e^{-\theta_2 r_2 - k_2}\right]. \tag{81}$$

This must hold for all r_1. Therefore, $\theta_{21} = \theta_1$.

Next, consider total utility U:

$$\begin{aligned} U &= \ln c_1 + \ln c_2 - g(a_1) - g(a_2) \\ &= 2\theta_1 r_1 + \theta_2 r_2 - g(a_1) - g(a_2) + 2\kappa_1 + k_2. \end{aligned}$$

From (55), the two EF conditions are $E_1\left[\frac{\partial U}{\partial r_1}\right] \geq g'(a^*)$ and $E_2\left[\frac{\partial U}{\partial r_2}\right] \geq g'(a^*)$, that is,

$$2\theta_1 \geq g'(a^*) \text{ and } \theta_2 \geq g'(a^*).$$

Intuitively (and as can be proven), the EF constraints should bind, else the CEO is exposed to unnecessary risk. Combining the binding version of these constraints with (80) yields (79).

Step 2: Optimality of log-linear contracts

We next verify that optimal contracts should be log-linear. Equation (55) yields $d(\ln c_2)/dr_2 \geq g'(a^*)$. The cheapest contract involves this local EF condition binding, that is,

$$d(\ln c_2)/dr_2 = g'(a^*) \equiv \theta_2. \tag{82}$$

Integrating yields the contract

$$\ln c_2 = \theta_2 r_2 + B(r_1), \tag{83}$$

where $B(r_1)$ is a function of r_1, which we will determine shortly. It is the integration "constant" of equation (82) viewed from time 2.

We next apply the PS constraint (61) for $t = 1$:

$$1 = E_1\left[\frac{c_1}{c_2}\right] = E_1\left[\frac{c_1}{e^{\theta_2 r_2 + B(r_1)}}\right] = E_1\left[e^{-\theta_2 r_2}\right] c_1 e^{-B(r_1)}, \tag{84}$$

Hence, we obtain

$$\ln c_1 = B(r_1) + K', \tag{85}$$

where the constant K' is independent of r_1. (In this proof, K', K'' and K''' are constants independent of r_1 and r_2.) Total utility is

$$U = \ln c_1 + \ln c_2 + K'' = \theta_2 r_2 + 2B(r_1) + 2K' + K''. \tag{86}$$

We next apply (55) to (86) to yield $2B'(r_1) \geq g'(a^*)$. Again, the cheapest contract involves this condition binding, that is, $2B'(r_1) = g'(a^*)$. Integrating yields

$$B(r_1) = g'(a^*)\frac{r_1}{2} + K'''. \tag{87}$$

Combining (87) with (85) yields: $\ln c_1 = g'(a^*)\frac{r_1}{2} + \kappa_1$, for another constant κ_1. Combining

(87) with (83) yields:

$$\ln c_2 = g'(a^*)\left(\frac{r_1}{2} + r_2\right) + \kappa_1 + k_2,$$

for some constant k_2.

Comparative statics for fixed pay ϕ. We have

$$\frac{\partial \phi^*}{\partial \eta} = K_\eta(3b^4S^4 - b^2gwS^2\left(\eta\sigma^2 - 2S\right) + 2g^2\eta\sigma^2 S),$$

where

$$K_\eta = \frac{b^2\sigma^2S^2}{2\left(b^2S^2 + g\eta\sigma^2\right)^3}.$$

Hence, $\frac{\partial \phi^*}{\partial \eta} > 0$ if and only if

$$3b^4S^4 + 2b^2gS^3 + 2g^2\eta\sigma^2 S > b^2gS^2\eta\sigma^2.$$

Further, we have

$$\frac{\partial \phi^*}{\partial \sigma} = K_\sigma(3b^4S^4 - b^2gS^2\left(\eta\sigma^2 - 2S\right) + 2g^2\eta\sigma^2 S),$$

where

$$K_\sigma = \frac{b^2\eta\sigma S^2}{\left(b^2S^2 + g\eta\sigma^2\right)^3}.$$

Hence, $\frac{\partial \phi^*}{\partial \sigma} > 0$ if and only if

$$3b^4S^4 + 2b^2gS^3 + 2g^2\eta\sigma^2 S > b^2gS^2\eta\sigma^2.$$

We have

$$\frac{\partial \phi^*}{\partial g} = K_g(b^6S^6 + 3b^4g\eta\sigma^2S^4 - 2b^2g^2\eta\sigma^2S^2\left(\eta\sigma^2 - S\right) + 2g^3\eta^2\sigma^4 S),$$

where

$$K_g = \frac{b^2S^2}{2g^2\left(b^2S^2 + g\eta\sigma^2\right)^3}.$$

Hence, $\frac{\partial \phi^*}{\partial g} > 0$ if and only if

$$b^6S^6 + 3b^4g\eta\sigma^2S^4 + 2b^2g^2\eta\sigma^2S^3 + 2g^3\eta^2\sigma^4 S > 2b^2g^2\eta\sigma^2S^2\eta\sigma^2.$$

B. Further Detail on Holmstrom-Milgrom (1987)

This section provides further details on the role played by exponential utility, a pecuniary cost of effort, and Normal noise in the Holmstrom and Milgrom (1987) model. In the general objective function (10), we now assume $u(x) = x$ so that the cost of effort is non-pecuniary, and have a general (rather than exponential) function $v(\cdot)$. We assume the contract can be rewritten $c(V) = \phi + z(V)$ where ϕ is the fixed component of the contract and $z(V)$ is a possibly non-linear function; this is without loss of generality since ϕ can be zero. The agent's first-order condition becomes

$$\mathrm{E}\left[v'\left(\phi + y\left(S + b\left(S\right)a + \varepsilon\right)\right) y'\left(V\right) b\left(S\right)\right] = g'\left(a\right).$$

The agent's reservation utility w affects the fixed salary ϕ, which in turn has two effects on his effort choice. First, it affects his benefit from effort. A higher ϕ reduces the marginal utility of money $v'(\phi + z(V))$ because the agent is risk averse. However, it does not affect the marginal cost of effort, because effort entails disutility rather than a financial expenditure. Thus, with a linear contract, the optimal action will depend on ϕ. Second, it affects the agent's attitude towards risk ε. The noise realization affects the agent's benefit from effort since he is risk-averse. For example, if noise turns out to be high, then the agent will be highly-paid even with low effort; thus, the benefits from working are lower: $v'(\cdot)$ falls. Hence, the agent will integrate over the possible noise realizations when making his effort choice. Since ϕ also lies in the marginal felicity function $v'(\cdot)$, it affects the agent's attitude towards risk and thus his effort choice.

To remove the first effect, HM assume a pecuniary cost of effort, which corresponds to $v(c) = c$ and a general $u(\cdot)$ in the objective function (10). Thus, the first-order condition becomes

$$\mathrm{E}\left[u'\left(\phi + z\left(S + b\left(S\right)a + \varepsilon\right)\right)\left(z'\left(V\right) b\left(S\right) - g'\left(a\right)\right)\right] = 0.$$

Now, the marginal benefit of effort $z'(V)\,b(S)$ and the marginal cost of effort $g'(a)$ are on the same footing: both lie inside the final term in parentheses. A high fixed wage ϕ reduces the benefit of effort, but also the cost of effort because this cost is in financial terms. However, ϕ still affects the attitude to risk since it is inside the $u'(\cdot)$ term. Thus, to remove this second effect, we also need exponential utility, $u(x) = -e^{-\eta x}$, so that the objective function (10) becomes

$$\mathrm{E}\left[-e^{-\eta(\phi + z(V) - g(a))}\right] = e^{-\eta\phi}\,\mathrm{E}\left[-e^{-\eta(z(V) - g(a))}\right]$$

with first-order condition

$$\mathrm{E}\left[\eta e^{-\eta(z(S+b(S)a+\varepsilon)-g(a))}\left(z'\left(S + b\left(S\right)a + \varepsilon\right) b\left(S\right) - g'\left(a\right)\right)\right] = 0. \tag{88}$$

The fixed salary ϕ, and thus the reservation utility w, is now irrelevant. However, the contract is still very difficult to solve as we cannot factor out the noise ε. Again, since the incentive constraint (88) must hold only on average, there are many possible contracts that will implement a given action a^*. The contract will depend critically on the distribution of noise ε, which poses important practical challenges as the noise distribution is often unknown. Only with Normal noise are we able to calculate the expectation, since then $\mathrm{E}\left[e^{-\eta\varepsilon}\right] = \frac{1}{2}\eta^2\sigma^2$.