Executive Pay, Career Path and Firm Size*

Jaeyoung Sung† and Peter L. Swan‡

University of Illinois at Chicago
and
Australian School of Business, University of New South Wales

This version: May 8, 2008
Preliminary: comments welcome

Abstract

In this paper we provide a simple agency model of executive pay as it relates to both firm size and executive career concerns as managers are recruited by firms of different sizes over their careers. Unlike the matching literature which assumes that the talent of managers is common knowledge, we explicitly model the hiring decisions of firms of different sizes when managerial talent is unknown but reflected imperfectly in firm performance in the manager's early career. As a result our agents' reservation utility levels are endogenously determined as outcomes of labor market equilibrium. Our model, together with new empirical evidence, explains the well-known but until now inexplicable finding that pay-performance sensitivity diminishes with size.

Our empirical results for 3,238 CEO stints show that a doubling of firm size raises CEO effort productivity by about 52% and ability productivity by about 108%, whereas pay also increases by about 53%, even after controlling for ability. Results are similar for accounting measures of performance, as well as market-based. The variation in ability for CEOs of large firms is less that for small firms but there is still considerable variation. This is unlike recent findings which show negligible variation in ability with a matching model that assumes *ex ante* known ability. A very high proportion of CEOs (26%) display negative career productivity but appear to suffer a negligible pay penalty unlike performers who are rewarded. However, unlike negative performers whose career stint length pay elasticity is a negative 13%, positive performers gain a 12% premium with respect to stint length and enjoy a 24% longer CEO stint.

Key words: executive pay, firm size, career concern, CEO talent, principal-agent, optimal contract.
JEL Classification: G34, J41, J44, L25

* We wish to thank the Australian Research Council (ARC) for financial support and Vijay Philip, Jasper Timm and Cybele Wong for valuable programming and related support. We thank David Feldman and Denzil Fiebig for useful comments.

† Contact details: Department of Finance, College of Business Administration, University of Illinois at Chicago, email: jsung@uic.edu; Tel: (312) 996-0720

‡ Contact details: Department of Banking and Finance, Australian School of Business, University of New South Wales, Email: peter.swan@unsw.edu.au; Tel: 61 (0)2 9385 5871.

1. Introduction

One of the best established facts in the social sciences over the last 50 years is that pay levels executives increase in the size of the organization with approximately a one-third higher pay level for each doubling of firm size. The strong positive correlation between firm size and executive pay has become one of the most highly documented facts in the area of executive compensation, over both many decades and numerous countries.[1] In this paper we provide an agency theory explaining the career paths of managers employed by firms of different sizes that are subject to moral hazard due to the unobservable nature of their actions and potentially self-interested motivations. We estimate the stochastic CEO career production function in terms of total assets for CEOs subject to moral hazard for S&P 1500 firms over the period, 1992-2005, based on both stock market and accounting career productivity measures. For the entire sample of 3,238 market and accounting CEO-stint productivity measures, we find that CEO productivity in real terms is increasing at the rate of approximately 52 percent in estimated CEO effort for each doubling of firm size for a manager of given talent, providing a rationale for numerous pay-scale findings. Similar results are obtained for accounting measures of performance, as well as market-based measures. Remarkably, after controlling for managerial talent, incentive

---

[1] See, for example, Roberts (1956), Lewellen and Huntsman (1970), Cosh (1975), Murphy (1985), Baker, Jensen, and Murphy (1988), Kostiuk (1990), Barro and Barro (1990), Rosen (1992), Joskow et al (1993), Rose and Shepard (1997), and Frydman and Saks (2007). Their findings suggest that a 100 percent larger firm will pay its CEO about thirty-three percent more. Zhou (1999) finds a positive correlation between executive pay and firm size for Canadian firms. Kaplan (1997), Kato (1997) and Kato and Kubo (2006) find a similar result for Japanese firms; Cosh and Hughes (1997) and Conyon (1997), McKnight, and Tomkins (1999) and Conyon and Murphy (2000) for British firms; and Izan *et al* (1998) and Merhabi, Pattenden, Swan and Zhou (2006) for Australian firms.

contract and a variety of other influences, pay increases at almost precisely the same rate (53%) for each doubling of assets under management.

CEO productivity is also increasing at approximately 108 percent in estimated CEO talent which is scaled in (real) dollar terms along with all our other measures. Thus ability productivity increases at the rate of 54 percent greater than effort productivity with scale, increasing to about 71 percent faster for small firms since effort productivity is less scale sensitive in small firms. Managers of companies need to be smart rather than work hard and this is especially so for large companies. This not only provides large firms with a strong desire to attract the most talented managers, but also indicates that ability or talent, unlike CEO effort, raises the productivity of the entire organization. Moreover, a more talented CEO captures much of the increase in market or accounting productivity that he generates. Based on the market measure for the entire sample he gains 67% of his ability factor but is negligible for CEOs with a negative ability factor. Large firms pay an ability premium at the rate of 136%, falling to 40% for small firms. With the accounting-based performance measure the ability pay factor increases at the rate of 265% for large firms but is negligible for both small firms and negative performers. One possible explanation for the lack of significant penalties for poor performance may be the desire to encourage CEO risk-taking as the CEOs focus on the firm would indicate that he is likely to be less diversified than shareholders. Section 2 reviews the literature, the model is developed in Section 3 and the empirical estimation is in Section 4. Section 5 concludes.

2. Literature Review

Recently, Gabaix and Landier (2007) develop a calculable neoclassical competitive equilibrium model of CEO pay in which the talents of CEOs are matched to firms of different sizes. Talents are presumed to be known to the market even in advance of performance. The model predicts that the six-fold increase in U.S. CEO pay between 1980 and 2003 is due to a six-fold increase in market capitalization of large U.S. companies over the same period. Based on the careers of 3,238 CEOs we find that, controlling for talent, CEO pay increases by 53 percent for each doubling in size, so that their simulated results differ from our empirical findings. Moreover, and quite remarkably, a negligible difference in managerial talent of only 0.016% between the CEO ranked number 250 and the top CEO accounts for pay for the top manager that is 500% higher than for manager number 250. These Gabaix and Landier findings immediately raise the quandary as to why the market for executive talent does not clear. In particular, if the alternative for the most talented executive assigned to the largest firm is to be employed by a smaller firm, say # 250, why does not the largest firm offer (say) just $1 more than firm #250 for a manager of almost precisely the same ability, rather than pay 500% more? Unlike Gabaix and Landier we explicitly model the clearing of the market for talent.

Contrary to the Gabaix and Landier findings indicating practically zero volatility, we find that the volatility of CEO ability is significant and, for the large firms investigated by Gabaix and Landier (2007) lower that of smaller firms. In fact, our findings are consistent with our model of the clearing of the market for talent. Gabaix and Landier (2007) find that over time (as opposed to the cross section) CEO pay moves proportionately to firm market capitalization. In effect, they find that the marginal impact of managerial talent increases in proportion to the value of the firm. We confirm the correctness of this approach with reference to talent in isolation with the estimated scale elasticities for CEO talent ranging from 108 to 116 percent. Kaplan and Rauh (2007) also present evidence consistent with the scale economy and technological explanations underlying the implications of Gabaix and Landier (2007) and Rosen (1982). Kaplan, Klebanov and Sorensen (2007) find that it is possible to measure CEO skills with success being more strongly related to execution rather than team skills.

In contrast to Gabaix and Landier (2007), Edmans, Gabaix, and Landier (2007) and our own findings, Frydman and Saks (2007) find only a weak relationship between compensation and firm size from the late 1940s to the mid-1970s. These findings for the earlier period suggest that technological advances in the last 35 years have increased the ability of able executives to manage very large companies successfully.

The pay sensitivity issue has emerged as an important issue in the agency literature since Jensen and Murphy (1990) argued that there is a negative relationship between sensitivity and firm size. However their claim is not fully satisfactory in that their argument is based on a standard agency model without taking the firm size effect into account. Our model explicitly accounts for firm size effects in contracting and our empirical estimates show that pay-performance sensitivity optimally falls with size. Moreover, our modeling tells us how sensitivities are affected by executives' career paths. Given our estimated elasticities, our model predicts that a manager recruited from a small to a large firm will be given a lower pay-performance sensitivity that a manager recruited from a firm of the same size. This is because the large firm's performance provides a more reliable signal of the manager's ability with less risk being borne by the CEO and thus warrants the use of higher-powered incentives. We find some support for this hypothesis based on the careers of 76 CEOs who move from one firm to another.

Baker and Hall (2004) estimate a form of a production function for CEO output based on CEOs being of the same ability. CEO effort increases in pay-performance sensitivity of the manager but there is no independent impact of managerial talent or scale effects of talent. Their estimated effort elasticity based on market value ranges from as low as 37 percent up to 66 percent and thus overlaps with our estimate for the entire sample based on market performance of 52 percent.

Our third focus, apart from the pay-size relationship and size-sensitivity is executive career concern. Fama (1980) provocatively argued in the absence of formal modeling that the managerial labor market could provide a perfect substitute for incentive pay by rewarding managers with high reputations for talent even though there is a moral hazard problem due to the unobserved nature of the manager's actions. Holmstrom (1999) formally modeled such career concern issues to show that the less is known about managerial ability the greater the incentive for the manager to supply effort as early on in the manager's career as it hard for the market to distinguish between effort and talent which become substitutes within his model. While there are limited circumstances in which Fama's confidence in the market as an antidote to moral hazard issues is correct, risk aversion and discounting qualify his conclusions. Gibbons and Murphy (1992) also model career concern issues to argue that explicit contracts should provide stronger incentives as executives approach retirement as the impact of the implicit incentives provided by the labor market decline with the prospect of retirement. Chung and Esö (2007) examine a sequential signaling equilibrium in which the manager may choose not to reveal the most informative public signal as to his ability.

In a neoclassical world, it can be socially optimal that large firms (or firms with better production technologies) hire managers with higher abilities. Thus, it is conventionally argued that there should be a positive relationship between pay and firm size because large firms hire high ability managers who deserve higher pay.

However, in an agency world, a large firm (with better production technology) may not always be willing to hire a high-ability manager because he may demand too high a salary. We argue that even when a large firm hires a low-ability manager, the expected pay for the low ability manager can be higher than that for a high-ability manager who is hired by a small firm, if the large firm's productivity and the firm size are sufficiently higher and larger than those of the small firm.

Note that our modeling shows that the executive pay-size relationship is critically affected at least by the following two components: the amount of "ability scarcity" rents each manager enjoys as size varies, and the structure of the managerial incentive scheme. Understanding of the first component, not to mention the second, is also important when we try to compare executive pay sizes across firms, because the way the managerial rent is allocated by small firms can be very different from that by large firms. Thus, it should be clear that an economic model that can explain executive size differential across firms should contain rent-allocational issues as well as moral hazard issues.

In order to understand the methodology we propose, it is necessary to describe the foundations of our approach. In approximate terms the structure of executive pay consists of two parts, non-performance-based and performance-based compensation. An integral part of non-performance-based compensation is the agent's reservation utility, which is in turn determined by the agent's labor market opportunities. As a result, the pay-size relationship can be significantly affected by the agent's reservation utility levels. Since agents in the labor market for large firms may demand higher reservation utility levels than agents in the labor market for small firms because market opportunities are better, the role of reservation utility levels can be critical, particularly when pay levels are compared across firm sizes. How then are agents' reservation utilities determined in the labor market?

Since the existing agency literature typically assumes the agent's reservation utility as an exogenous parameter, it does not provide a model that can be used to compare different executive pay levels across firms with different sizes. We build a principal-agent model with reservation utility levels endogenously determined though labor market competition. We believe our agency model with enodgenized labor market competition enables us to shed light on how the executive pay level changes as executives move from one firm to another.

The model we build is a multi-period contracting principal–agent model with career concerns. It is constructed by extending continuous-time agency models developed by Holmstrom and Milgrom (1987) and Schättler and Sung (1993) with some insights from Zhou and Swan (2003). In particular, there are two firms, one small and the other, large. For each continuous time period, each firm signs a contract with one agent chosen from two candidates with unknown abilities. After the first continuous-time period, each firm again makes its hiring decision for the second continuous-time period. The hiring decision will be made with each agent's past performance record taken into account. His past record would provide each agent with differing negotiating power for second-period contracting, and thus a different reservation utility. Our model would allow agents' reservation utility levels to be endogenously determined as the two firms compete to hire the better of the two managers and thus we would expect to see quite different pay outcomes depending on firm size and "manager reputation" based on track record.

The second-period contract sensitivity of the large firm can be lower or higher than that of the small firm, depending on the relative productivity of the two firms. If per-dollar volatilities of firms' profits are the same across firms, and if there are constant (decreasing, increasing) expected returns to scale across firms, contract sensitivities to realized profits are the same across firms (are decreasing, increasing with firm size), even when total dollar volatilities are increasing with the firm size.

It also has implications for managerial career paths which can also affect the contract sensitivity. It depends on how managerial ability contribution to the total profit changes as the firm size increases. Suppose that the managerial ability contribution is independent of the firm size. Then we expect that a manager who previously worked for a large firm will be given a contract with a lower sensitivity than a manager who worked for a small firm. Suppose that the managerial ability contribution increases faster than the total profit volatility, as the firm size increases. Then we expect that a manager who previously worked for a large firm will be given a contract with a higher sensitivity than a manager who previously worked for a small firm. Thus our model has a number of clear-cut testable implications.

In the existing framework, e.g., Rosen (1982), talent is known and there is no explicit modeling of hiring decisions or how the participation constraint and reservation utility level of executives is determined. Nor is there any incentive–based compensation in place so that these models have nothing to say about the relationship between pay-performance sensitivity and firm size. Moreover, being purely competitive models of the labor market, there is no possibility of rent extraction by the CEO.

There have been several other attempts to try and reconcile the pay-size premium with possible explanations of the phenomena put down to effects, such as, compensating differentials by Dunn (1986), union status by Lewis (1986) and efficiency wages by Krueger and Summers (1988). However, even when these factors are taken into consideration a large proportion of the disparity in pay still remains. Ferrall (1997) extends the original Rosen (1982) model to consider two skills possessed by engineers that vary in usefulness according to the position in the hierarchy. He finds that better engineers tend to select into higher paid jobs but unobserved skills accounts for most of the wage differential. Idson and Oi (1999) and Bayard and Troske (1999) both find that workers in larger firms achieve higher labor productivity. However, given that existing measures of labor productivity are all endogenously determined by firm size this fact cannot be conclusively proven. A study by Bebchuk and Grinstien (2006) finds that that there is an economically and statistically significant positive relationship between CEO compensation and the CEO's past decisions to increase firm size, by means of increasing the number of shares on issue.

3. The Two-Period Career Concerns Model

There are two firms, S (small) and L (large), and two agents a and b with unknown abilities θ^a and θ^b, respectively. We assume that $\theta^{\hat{a}}$, for $\hat{a} = a,b$, is normally distributed with a mean of $m_{\theta^{\hat{a}}}$ and a variance of σ_θ^2, and that θ^a and θ^b are independent of each other . The firms are risk neutral and the agents are risk averse with a constant absolute risk aversion (CARA) coefficient of r.

There are two periods with three dates, 0, 1, and 2. Contracting between the two firms and two agents occurs at time 0 and 1. At time $t-1$, for $t = 1,2$, firm i $(= S, L)$ hires agent $\hat{a}$ $(= a,b)$, and the agent exerts effort $e_{t-1}^{\hat{a}}$ to produce outcome $Y_t^i(\hat{a})$, where

$$Y_t^i(\hat{a}) = e^{\hat{a}} f(K^i) + \theta^{\hat{a}} g(K^i) + h(K^i)\varepsilon_t^i, \tag{1}$$

and $K^S < K^L$; and ε_t^i , for $t = 1,2$ and $i = S, L$, are independent standard-normal random variables, each of which are distributed with a mean of zero and a standard deviation of σ .[2] The ability $\theta^{\hat{a}}$ may represent agent $\hat{a}$'s decision-making competency or information-gathering ability to identify better investment opportunities. The functions $f(K^i)$, $g(K^i)$ and $h(K^i)$, respectively, describe how firm size and scale economies affect the agent's marginal productivity of effort, ability and the way risk (dollar volatility) varies with firm size. For effort $e^{\hat{a}}$, agent $\hat{a}$ incurs a personal (monetary) cost of $c(e) = (\kappa / 2)e^2$ for some $\kappa > 0$. We also assume that $f(K) = K^{\gamma_f}$, $g(K) = K^{\gamma_g}$, and $h(K) = K^{\gamma_h}$, for some $\gamma_f, \gamma_g, \gamma_h > 0$. Then, Eq. (1) implies that the expected outcome is given in the form of a stochastic Cobb-Douglas production function of labor effort e and capital K in additive form.

At time $t(=1,2)$, agent $\hat{a}$ working for firm i is compensated by an amount $S_t^i(Y_t^i(\hat{a}))$, and the utility of agent $\hat{a}$ working for firm i takes the form $-\exp\left(-r\sum_{t=1}^{2}\left\{S_t^i(\hat{a}) - c(e_{t-1}^{\hat{a}})\right\}\right)$. Without loss of generality, we assume, for $t = 1,2$, that there is a linear pay schedule with a fixed and a share component:

$$S_t^i(Y_t^i(\hat{a})) = \alpha_{t-1}^i(\hat{a}) + \beta_{t-1}^i(\hat{a})Y_t^i(\hat{a}) .$$

At time 0, agent $\hat{a}$'s reservation utility is $-\exp\left(-rW_0^{\hat{a}}\right)$, where $W_0^{\hat{a}}$ is called the certainty equivalent reservation wealth level. At time 1, since outcomes of agents'

[2] This outcome only describes the production side of the firm: the larger the firm size, the higher the production. However, it is possible to discuss the issue of the optimal size of the firm without losing the essence of the economics of the problem. If so desired, one can (almost trivially) incorporate the issue into our model without affecting our results. For this purpose, one may explicitly introduce the cost side of the firm by assuming that the expected marginal cost of the small firm is increasing with firm size faster than that of the large firm, and also assuming for simplicity that the cost is independent of the agent's effort. Then, optimal sizes of both the small and large firm can be justified.

effort are realized, and provide better information about agents' abilities, firms compete for better agents, and as a consequence, the certainty equivalent wealth level of agent $\hat{a}$ who previously worked for firm k changes to $W_1^{\hat{a}}(Y_1^k(\hat{a}))$.

3.1 The Second-Period Contracting

In this section, contracting occurs twice: initial contracting at time 0, and re-contracting at time 1. That is, this dynamic contracting problem consists of the first and second-period contracting problems. We consider the second-period problem first so as to solve the overall problem recursively.

Suppose that at time 0, agent $\hat{a}$ worked for firm k $(=S, L)$ and at time 1, he is hired by firm i. The outcome of agent $\hat{a}$'s time-0 performance with firm k is realized at time 1 to be $Y_1^i(\hat{a})$, and firm i updates its belief on agent $\hat{a}$'s ability $\theta^{\hat{a}}$ based on the realized outcome $Y_1^i(\hat{a})$. Since both $Y_1^i(\hat{a})$ and $\theta^{\hat{a}}$ are normally distributed, and they are linearly related to each other through Eq. (1), $\theta^{\hat{a}}$ conditional on $Y_1^i(\hat{a})$ is normally distributed with its mean and variance given as follows:

$$E[\theta^{\hat{a}} \mid Y_1^k(\hat{a})] = m_{\theta^{\hat{a}}} + p^k\left(Y_1^k(j) - e_0^{\hat{a}} f(K^k) - m_{\theta^{\hat{a}}} g(K^k)\right), \tag{2}$$

$$p^k := \frac{\sigma_\theta^2 g(K^k)}{\sigma_\theta^2 g^2(K^k) + \sigma^2 h^2(K^k)}, \tag{3}$$

$$Var[\theta^{\hat{a}} \mid Y_1^k(j)] = \frac{\sigma_\theta^2 \sigma^2 h^2(K^k)}{\sigma_\theta^2 g^2(K^k) + \sigma^2 h^2(K^k)}. \tag{4}$$

That is, common beliefs on agent abilities are updated over time with their conditional means and variances given by Eq.'s (2) and (4).

As a result, the second-period contracting problem for firm i hiring agent $\hat{a}$ who worked for firm k for the first period can be stated as follows.

Problem 1 (The second-period contracting.) Choose pay contract $S_2^i(\hat{a})$ to maximize the expected profit to the shareholders in period 2 conditional on the agent's performance outcome in period 1:

$$E[Y_2^i(\hat{a}) - S_2^i(\hat{a}) \mid Y_1^k(\hat{a})], \text{ subject to}$$

(1) $Y_2^i(\hat{a}) = e_1^{\hat{a}}(i) f(K^i) + \theta^{\hat{a}} g(K^i) + h(K^i)\varepsilon_2^i$,

$$Y_1^k(\hat{a}) = e_0^{\hat{a}}(k) f(K^k) + \theta^{\hat{a}} g(K^k) + h(K^k)\varepsilon_1^k,$$

(2) $e_1^{\hat{a}}(i) \in \arg\max_{\hat{e}} E\left[-\exp\left\{-r\left(S_2^i(\hat{a}) - c(\hat{e})\right)\right\} \mid Y_1^k(\hat{a})\right]$

$$\text{s.t.} \qquad Y_2^i(\hat{a}) = \hat{e} f(K^i) + \theta^{\hat{a}} g(K^i) + h(K^i)\varepsilon_2^i,$$

$$Y_1^k(\hat{a}) = e_0^{\hat{a}}(k) f(K^k) + \theta^{\hat{a}} g(K^k) + h(K^k)\varepsilon_1^k,$$

(3) $E\left[-\exp\left\{-r\left(S_2^i(\hat{a}) - c(e_1^{\hat{a}}(i))\right)\right\} \mid Y_1^k(\hat{a})\right] \geq -\exp\left(-rW_1^{\hat{a}}\right)$.

The first constraint is simply the production function in period's 1 and 2. The second constraint is the agent's effort incentive constraint conditional on the outcome in period 1, and the third constraint is the participation constraint given the agent's reservation utility in period 1. The first order condition (FOC) from the incentive constraint combined with the participation constraint implies that the second period pay schedule:

$$S_2^i(\hat{a}) = \underbrace{W_1^{\hat{a}} + c(e_1^{\hat{a}}) + \frac{r}{2}\left(\frac{c_e(e_1^{\hat{a}})}{f(K^i)}\right)^2\left(\frac{g^2(K^i)\sigma_\theta^2\sigma^2 h^2(K^k)}{\sigma_\theta^2 g^2(K^k)+\sigma^2 h^2(K^k)} + \sigma^2 h^2(K^i)\right)}_{\text{fixed compensation}}$$

$$+\underbrace{\left(\frac{c_e(e_1^{\hat{a}})}{f(K^i)}\right)\left\{Y_2^i(\hat{a}) - \left(e_1^{\hat{a}} f(K^i) + E[\theta^{\hat{a}} \mid Y_1^k] g(K^i)\right)\right\}}_{\text{performance-based compensation}}. \tag{5}$$

Note that the sensitivity of the contract, or the sensitivity of the compensation to the realized outcome $Y_2^i(\hat{a})$, is $\frac{c_e(e_1^{\hat{a}})}{f(K^i)}$. The structure of Eq. (5) is well-known, consisting of two parts: fixed and performance-based compensations.

The first term of the fixed compensation is the agent's certainty equivalent reservation wealth, the second the cost of effort, and the third the compensation-risk premium. The performance-based compensation is in proportion $\frac{c_e(e_1^{\hat{a}})}{f(K^i)}$ to a metric for unexpected outcome $Y_2^i - E[Y_2^i \mid Y_1^k]$ which is realized minus expected outcomes. The performance-based compensation constitutes a compensation risk to the agent, on which the agent demands a risk premium as much as the third term of the fixed compensation.

By Eq.(5), the expected profit of firm i at time 2 is

$$\pi^i(\hat{a}(k), W_1^{\hat{a}(k)}) := E[Y_2^i(\hat{a}) - S_2^i(\hat{a}) \mid Y_1^k(\hat{a})] = g(K^i)E[\theta^{\hat{a}} \mid Y_1^k(\hat{a})] + \Phi(K^k, K^i) - W_1^{\hat{a}},$$

where:

$$\Phi(K^k, K^i) = \max_e e(K^i)^{\gamma_f} - \frac{\kappa}{2}e^2 - \frac{r}{2}\left(\frac{\kappa e}{(K^i)^{\gamma_f}}\right)^2\left(\frac{(K^i)^{2\gamma_g}\sigma_\theta^2\sigma^2(K^k)^{2\gamma_h}}{\sigma_\theta^2(K^k)^{2\gamma_g} + \sigma^2(K^k)^{2\gamma_h}} + \sigma^2(K^i)^{2\gamma_h}\right). \tag{6}$$

Substituting the FOC with respect to effort e back into the RHS of Eq. (6), we have

$$\Phi(K^k, K^i) = \frac{1}{2} e(K^i)^{\gamma_f} = \frac{1}{2} \frac{1}{\kappa(K^i)^{-2\gamma_f} + r\kappa^2\sigma^2 \left(\frac{(K^i)^{2(\gamma_g - 2\gamma_f)} \sigma_\theta^2}{\sigma_\theta^2 (K^k)^{2(\gamma_g - \gamma_h)} + \sigma^2} + (K^i)^{2(\gamma_h - 2\gamma_f)} \right)}. \quad (7)$$

3.1.1 Pay Sensitivity and Firm Size

The pay sensitivity has become an important issue in the agency literature since Jensen and Murphy (1990) argued that there is empirically a negative relationship between the sensitivity and firm size. The firm i's problem in Eq.(6) enables us to relate the sensitivity to firm size.

The FOC for Eq. (6) also implies that for firm *i* hiring agent $\hat{a}$ who worked for firm *k*, the sensitivity of the contract for the second period to motivate the agent to exert effort is

$$\frac{\kappa \hat{e}}{(K^i)^{\gamma_f}} \equiv \beta^i(\hat{a}(k)) \equiv \beta^i = \frac{1}{1 + r\kappa\sigma^2 \left(\frac{(K^i)^{2(\gamma_g - \gamma_f)} \sigma_\theta^2}{\sigma_\theta^2 (K^k)^{2(\gamma_g - \gamma_h)} + \sigma^2} + (K^i)^{2(\gamma_h - \gamma_f)} \right)}. \quad (8)$$

Note that the sensitivity expression, Eq. (8), does not directly depend on agent type $\hat{a}$, because of our assumption that the risk aversion and effort cost functions for both agents are identical. The sensitivity, however, depends on the agent's work experience *k,* because the experience affects the volatility of the second-period outcome as the distribution of the agent ability level is updated by the work experience. Also note that the sensitivity is denoted by β^i for brevity, instead of $\beta^i(\hat{a}(k))$.

Eq. (8) immediately relates the sensitivity to the size of the firm as follows:

Proposition 1: *Suppose that* K^k *, the size of the firm for which the manager previously worked for is given. Then:*

(i) *The sensitivity is inversely related to the firm size, if either the relative scale elasticities,* $\gamma_h - \gamma_f \geq \gamma_f - \gamma_g > 0$ *or* $\gamma_g > \gamma_f > 0$ *and* $\gamma_h > \gamma_f > 0$.

(ii) *The sensitivity is positively related to the firm size, if either* $0 < \gamma_g - \gamma_f \leq \gamma_f - \gamma_h$ *or* $\gamma_f > \gamma_g > 0$ *and* $\gamma_f > \gamma_h > 0$.

Proofs are given in the *Technical Appendix.*

Proposition 1 suggests that, for example, the large firm offers a lower- (higher-) powered incentive contract than the small firm, when expected marginal effort-productivity $(K^i)^{\gamma_f}$ is sufficiently lower (higher) than expected ability-productivity $(K^i)^{\gamma_g}$ and volatility growth $(K^i)^{\gamma_h}$ over firm size. Substituting our empirically derived elasticities estimated in Tables 2 to 4 below, we find, based on stock market productivity measures for the entire sample, that $\gamma_g = 1.09 > \gamma_h = 0.86 > \gamma_f = 0.52$. The same inequalities are satisfied for accounting measures of productivity and for both large and small firms. Hence condition (i) rather condition (ii) is satisfied and pay-performance sensitivity is optimally negatively related to firm size.

This implication is in contrast with Baker and Hall (2004) who argued that the sensitivity is negatively related to the firm size because dollar volatilities of profits of large firms are higher than those of small firms. However, our Proposition 1 indicates that the relationship depends more on relative sizes of γ_f, γ_g and γ_h than it does on differences in dollar volatilities. For example, if $\gamma_f = \gamma_g = \gamma_h$, then sensitivities of both the large and small firms are identical, even though the total dollar volatility of the large firm can be way higher than that of the small firm.[3]

3.1.2 Career Path and Sensitivity

Eq. (8) also tells us how sensitivities are affected by executives' career paths.

Proposition 2: *If $\gamma_g > (<) \gamma_h$, then the second-period contract sensitivity for the agent who previously worked for the large firm is higher (lower) than the sensitivity for the agent who previously worked for the small firm.*

The statement is directly from Eq. (8). By Proposition 2, if $\gamma_g > \gamma_h$ as we find empirically, then we have $\beta^{LL} > \beta^{LS}$, and $\beta^{LS} > \beta^{SS}$; and if $\gamma_g < \gamma_h$, then $\beta^{LL} < \beta^{LS}$, and $\beta^{LS} < \beta^{SS}$, where L and S in the superscript denote the large and small firms, respectively, and the first superscript on β is the size of the firm the manager worked for in the first period, and the second superscript stands for the size of the firm the manager is currently working for.

To see this intuitively, note that Eq.(4) implies if $\gamma_g - \gamma_h > (<) \, 0$, the conditional variance of the agent ability given his performance with the first-period firm is inversely (positively) related to the firm size. That is, if $\gamma_g - \gamma_h > (<) \, 0$, the informativeness of the agent's past performance about his ability increases (decreases) with the firm size. Thus, if $\gamma_g - \gamma_h > (<) \, 0$, then the firm hiring a manager coming from the large firm would have lower (higher) outcome volatility and thus it provides

[3] Note the Baker and Hall (2004) do not take account of managerial ability at all insofar as ability is implicitly assumed to be identical for all managers. Moreover, the volatility growth $h(K) = \sigma K^{\gamma_h}$ is only implicit in Baker and Hall (2004) and is not formally modeled in their paper.

its manager with a higher-powered (lower-powered) contract than the other firm hiring a manager coming from the small firm would.

Since our empirical estimation set out in Tables 2 to 4 below shows that $\gamma_g - \gamma_h > 0$ for all market and accounting productivity measures, irrespective of whether all CEOs are included or the two sized-based samples, the model predicts that the contract sensitivity for the manager who moves from one large firm to another, or for that that matter stays in a large firm, will have higher contract sensitivity than the manager who moves from a small firm to a large firm. There is only a relatively small sample of 76 CEO movers. Of these 44, or 58 percent, have the predicted sign for beta.

3.2 Pay and Firm Size: Labor Market Equilibrium

In this section, we examine relationships between expected executive pay and firm size over the two contracting periods. As can be inferred from the form of the salary function in Eq.(5), the main issue in computing the expected executive pay is to understand how the executive reservation certainty-equivalent wealth level $W_1^{\hat{a}}$ is determined. It will be seen that the wealth level can depend on labor market competition for agents which is based on each agent's ability estimated from his past performance. In the labor market, each firm assesses each agent's ability given his past performance, and makes a job offer. Then, he chooses from job offers by the two firms. As a consequence, the agent's certainty reservation wealth is competitively determined.

For this, we model labor market competition between the two firms as follows. At time 0, agent a works for firm S, and agent b works for firm L. Then at time 1, there can be two possible cases: case (SS; LL) where agent a is rehired by firm S, and agent b is also rehired by firm L; and case (SL; LS) where agent a now works for firm L, and agent b now works for firm S.

Let agents $\hat{a}$ and $\hat{b}$ hired by firms S and L with certainty equivalent wealth levels being $W_1^{\hat{a}}$ and $W_1^{\hat{b}}$, respectively. We define the executive labor market equilibrium as follows. Each firm makes job offers to all agents on a first-come first-served basis. Each job offer made out to an agent by each firm is represented by a level of certainty equivalent wealth to the agent.

In particular, the two agents $(\hat{a}, \hat{b})$ receive job offers $(W_1^{\hat{a}S}, W_1^{\hat{b}S})$ from firm S and $(W_1^{\hat{a}L}, W_1^{\hat{b}L})$ from firm L. If agent $\hat{a}$ takes the offer by firm *S* before agent $\hat{b}$ does, then agent $\hat{a}$ enjoys a certainty equivalent wealth level of $W_1^{\hat{a}S}$, and agent $\hat{b}$ is hired by firm *L*.

It should be clear that each firm would like to hire an agent who would produce an expected profit to the firm at least as great as the other agent would. However, each firm's decision can also affect/be affected by the other firm's decision. We examine the following type of executive labor market equilibrium.

Definition 1: *The executive job market is in equilibrium with agents $\hat{a}$ and $\hat{b}$ choosing to work for firms S and L, respectively, if job offers $(W_1^{\hat{a}i}, W_1^{\hat{b}i})$ to agents $(\hat{a}, \hat{b})$ by firm i, for $i = S$ and L, satisfy the following properties.*

(i) (Profit maximization.)

$$W_1^{\hat{a}S} \in \arg\max_W \pi^S(\hat{a}, W) \ s.t. W \geq W_1^{\hat{a}L}, \ and \ \pi^S(\hat{a}, W_1^{\hat{a}S}) \geq \max_W \pi^S(\hat{b}, W) \text{ s.t. } W \geq W_1^{\hat{b}L}.$$

$$W_1^{\hat{b}L} \in \arg\max_W \pi^L(\hat{b}, W) \ s.t. W \geq W_1^{\hat{b}S}, \ and \ \pi^L(\hat{b}, W_1^{\hat{b}L}) \geq \max_W \pi^L(\hat{a}, W) \text{ s.t. } W \geq W_1^{\hat{a}S}.$$

(ii) (Expected zero profit condition for the small firm.)

$$\pi^S(\hat{a}, W_1^{\hat{a}S}) = \pi^S(\hat{b}, W_1^{\hat{b}L}) = 0.$$

Condition (i) implies that each firm chooses an agent to maximize its expected profit. In particular, given the offers by the large (small) firm to both agents, the small (large) firm finds itself to be better off improving the offer to agent $\hat{a}$ ($\hat{b}$) than it does to agent $\hat{b}(\hat{a})$. By this condition, we have $W_1^{\hat{a}S} = W_1^{\hat{a}L}$, $W_1^{\hat{b}L} = W_1^{\hat{b}S}$, $\pi^S(\hat{a}, W_1^{\hat{a}S}) \geq \pi^S(\hat{b}, W_1^{\hat{b}L})$ and $\pi^L(\hat{b}, W_1^{\hat{b}L}) \geq \pi^L(\hat{a}, W_1^{\hat{a}S})$. Condition (ii) suggests that agents' reservation certainty-equivalent wealth levels are determined by their job opportunities with the small firm, and that the small firm's expected profit is always driven to zero (perhaps by job/product market competition).

The next proposition sheds some light on equilibrium hiring decisions in the second period.

First, let us define:

$$\begin{aligned} A(K^S, K^L) &= \frac{1}{g(K^L) - g(K^S)} \left\{ \left(\Phi(K^L, K^L) - \Phi(K^L, K^S) \right) - \left(\Phi(K^S, K^L) - \Phi(K^S, K^S) \right) \right\}, \\ &= \frac{1}{g(K^L) - g(K^S)} \int_{K^S}^{K^L} \int_{K^S}^{K^L} \Phi_{K^k K^i}(K^k, K^i) dK^i dK^k. \end{aligned}$$

Then $A(K^S, K^L)\left(g(K^L) - g(K^S)\right)$ measures the comparative advantage of agent b over agent a in terms of marginal effort contribution to the large firm's expected profit over that of the small firm. If $A(K^S, K^L)\left(g(K^L) - g(K^S)\right)$ is positive, agent b's marginal effort-contribution to the expected profit of the large firm is relatively larger than that of agent a.

Proposition 3: *If $E[\theta^a - \theta^b \mid Y_1] \leq A(K^S, K^L)$, then in equilibrium, agents a and b are rehired by their firms, S and L, respectively, and*

$W_1^{aS} = g(K^S) E[\theta^a \mid Y_1^S(a)] + \Phi(K^S, K^S)$, and $W_1^{bL} = g(K^S) E[\theta^b \mid Y_1^L(b)] + \Phi(K^L, K^S)$.

If $E[\theta^a - \theta^b \mid Y_1] > A(K^S, K^L)$*, then in equilibrium, agents a and b are hired by firms,* L *and* S*, respectively, and agent expected wealth levels are*

$W_1^{bS} = g(K^S)E[\theta^b \mid Y_1^L(b)] + \Phi(K^L, K^S),$ *and* $W_1^{aL} = g(K^S)E[\theta^a \mid Y_1^S(a)] + \Phi(K^S, K^S).$

Remark 1: Note that regardless of the second period job mobility, individual agents' second-period certainty equivalent wealth levels remain unchanged in structure. For example, whether agent *a* (*b*) works for firm *S* or *L*, his second period certainty equivalent is given by $W_1^a = g(K^S)E[\theta^a \mid Y_1^S(a)] + \Phi(K^S, K^S)$ ($W_1^b = g(K^S)E[\theta^b \mid Y_1^L(b)] + \Phi(K^L, K^S)$).

Remark 2: Note that if $E[\theta^a - \theta^b \mid Y_1] \le A(K^S, K^L)$, then:

$$W_1^{bL} = W_1^{aS} + g(K^S)E[\theta^b - \theta^a \mid Y_1(b)] + \Phi(K^L, K^S) - \Phi(K^S, K^S);$$

and that if $E[\theta^a - \theta^b \mid Y_1] > A(K^S, K^L)$, then

$$W_1^{aL} = W_1^{bS} + g(K^S)E[\theta^a - \theta^b \mid Y_1(a)] + \Phi(K^S, K^S) - \Phi(K^L, K^S).$$

Remark 3: Note that $E[\theta^a - \theta^b \mid Y_1]\left(g(K^L) - g(K^S)\right)$ represents the marginal difference between agent *a* and *b* in terms of their marginal ability contribution to production with respect to firm size.

The small firm manager moves to the large firm in the second period if and only if his expected ability conditional on his first period performance, $E[\theta^a \mid Y_1]$, turns out to be sufficiently large, such that $E[\theta^a \mid Y_1] > E[\theta^b \mid Y_1] + A(K^S, K^L)$. In this sense, one may view the function *A* as a measure of executive job mobility: a high *A* means a low probability for small-firm managers to move to large firms. Unlike the matching literature, agents hiring/moving decisions are based not only on perceived ability levels but also on the volatility of their ability levels due to uncertainty as to what their ability really is. Hence their effort contributions are also influence by ability and its distribution.

Note that $E[\theta^b - \theta^a \mid Y_1] + A(K^S, K^L)$ measures the comparative advantage of agent b over agent a in terms of contributions by both ability and effort to the expected profit of the large firm. Thus, the agent worked for the small firm can be hired by the large firm only when his expected ability level is large enough to get over the large firm manager's comparative advantage.

3.3 The First-Period Contracting Problem

Agent $\hat{a}$'s, $\hat{a} = a, b$, effort choice decision for the first period can be affected by his job market prospects for the second period.

Problem 2: Choose the initial period pay schedule $S_1^i(\hat{a})$ to maximize expected first-period shareholder profit:

$$E[Y_1^i(\hat{a}) - S_1^i(\hat{a})]\text{, subject to}$$

(1) $Y_1^i(\hat{a}) = e_0^{\hat{a}} f(K^i) + \theta^{\hat{a}} g(K^i) + h(K^i)\varepsilon_1^i$,

(2) $e_1^{\hat{a}} \in \arg\max_{\hat{e}} E\left[-\exp\left\{-r\left(S_1^i(\hat{a}) - c(\hat{e}) + W_1^{\hat{a}}\right)\right\}\right]$,

$$\text{s.t.} \qquad Y_1^i(\hat{a}) = \hat{e} f(K^i) + \theta^{\hat{a}} g(K^i) + h(K^i)\varepsilon_1^i,$$

(3) $E\left[-\exp\left\{-r\left(S_1^i(j) - c(e_0^{\hat{a}}) + W_1^{\hat{a}}\right)\right\}\right] \geq -\exp\left(-rW_0^{\hat{a}}\right)$.

The main difference between Problems 1 and 2 is that in Problem 1, the agent's first-period wealth consists of not only $S_1^i(\hat{a}) - c(e_0^{\hat{a}})$, direct compensation from the firm net of effort cost, but also $W_1^{\hat{a}}$, the certainty equivalent wealth the agent can expect from the second period contracting. Young agents have career concerns that impact on their choice of their first managerial position.

Without loss of generality, we again assume optimal contracts are linear such that $S_1^i(\hat{a}) = \alpha^i + \beta^i Y_1^i(\hat{a})$ for agent $\hat{a}$ working for firm i.

Proposition 4: *Let firm i (=S,L) hires agent $\hat{a}$ (=a,b) at time zero. Then fixed and incentive parameters (α^i, β^i) for the optimal contract are given as follows:*

$$\begin{aligned}\alpha^{i\hat{a}} &= W_0^{\hat{a}} - \Phi(K^i, K^S) + c(\hat{e}^{\hat{a}}) - g(K^S)m_{\theta^{\hat{a}}} - \beta^i(\hat{e}^{\hat{a}} f(K^i) + m_{\theta^{\hat{a}}} g(K^i)) \\ &\quad + \frac{r}{2}(\beta^i + g(K^S)p^i)^2 \left(\sigma_\theta^2 g^2(K^i) + \sigma^2 h^2(K^i)\right).\end{aligned} \tag{9}$$

$$\begin{aligned}\beta^{i\hat{a}} &= \frac{c'(e)}{f^i} - g(K^S)p^i \\ &= \frac{1}{1 + r\kappa\left(K^i\right)^{-2\gamma_f}\left(\sigma_\theta^2 (K^i)^{2\gamma_g} + \sigma^2 (K^i)^{2\gamma_h}\right)} - \frac{\sigma_\theta^2 (K^i K^S)^{\gamma_g}}{\sigma_\theta^2 (K^i)^{2\gamma_g} + \sigma^2 (K^i)^{2\gamma_h}}\end{aligned} \tag{10}$$

Thus, the expected profit of firm i hiring agent $\hat{a}$ is

$$E[Y_1^i - S^i] = m_{\theta^{\hat{a}}}\left(g(K^i) + g(K^S)\right) - W_0^{\hat{a}} + \Phi(K^i, K^S) + \Psi\left(K^i\right), \tag{11}$$

where initial certainty equivalent wealth:

$$W_0^j = 2m_{\theta^j}(K^S)^{\gamma_g} + \Psi(K^S) + \Phi(K^S, K^S),$$

and

$$\Psi\left(K^i\right) = \frac{1}{2}\frac{1}{\kappa(K^i)^{-2\gamma_f} + r\kappa^2\left(K^i\right)^{-4\gamma_f}\left(\sigma_\theta^2 (K^i)^{2\gamma_g} + \sigma^2 (K^i)^{2\gamma_h}\right)}.$$

Recall that in the second period, there is no future career concern problems and the contract sensitivity is the marginal cost of effort per unit of marginal expected output,

$\frac{c'(e)}{f^i}$. However, Proposition 4 also implies that, in the first-period contracting, the sensitivity is adjusted for the agent's career concern by $g(K^S)p^i$. That is, in the first period, the contract sensitivity does not have to be equal to the marginal cost of effort per marginal expected output, because the agent has already built-in (implicit) incentives to work even without an explicit incentive contract. This kind of adjustment is well-known. See Gibbons and Murphy (1992).

Proposition 4 also implies that the expected pay differential between large and small firms is made up of the following components:

$$E[S_1^L]-E[S_1^S]=\underbrace{(m_{\theta^b}-m_{\theta^a})g(K^S)}_{\text{ability differential}}+\underbrace{\Psi(K^L)-\Psi(K^S)}_{\text{effort production differential}}+\underbrace{\Phi(K^S,K^S)-\Phi(K^L,K^S)}_{\substack{\text{compensation for disadvantage}\\ \text{in future labor market}}}.$$

There are three sources of the difference in pay size between executives of large and small firms: (1) the ability differential, (2) the effort production differential (due to difference in production functions between large and small firms), and (3) compensation for disadvantages the large-firm executive may experience in future executive labor markets. By contrast, in Edmans, Gabaix, and Landier (2007) pay differentials are entirely determined by talent/ability differentials. In this paper, disadvantages may come from the fact that the volatility of updated expectation of executive ability level after the first period will be higher for the large-firm executive, because dollar-return from production is more volatile for the large firm than it is for the small firm. That is, the large-firm profit outcome in the initial period provides a weaker signal as to agent ability than for the small-firm agent due to the volatility difference. This result is consistent with the signaling equilibrium of Chung and Esö (2007) in which an agent may choose to provide a weaker signal in the initial period.

Now, we examine effects of firm size on managerial salaries:

Proposition 5: Suppose that *at time 0, agents a and b are hired by firms S and L, respectively. Then the expected ability of the large-firm agent is at least as great as that of the small,* $m_{\theta^b} \geq m_{\theta^a}$. *Moreover, if* $\max[\gamma_g, \gamma_h] < 2\gamma_f$, *and* $\gamma_g < \gamma_h$, *then the first-period expected pay of the large firm manager is higher than that of the small firm.*

Evaluating the inequalities included in Proposition 5 utilizing the estimated elasticity values obtained in Table 4 below, the first inequality is satisfied but the second one is not as $\gamma_g > \gamma_h$. Hence in the first period of the manager's career, we cannot guarantee that the larger firm manager will be paid more in equilibrium than the smaller firm manager.

Remark: Since the two agents share the same the volatility of agent ability at time zero, the two firms only look at the mean of agent ability in their hiring decision at time zero. However, as already seen in Proposition 3, at time 1, both the (conditional) mean and volatility of agent ability are different between the two agents. Thus, hiring decisions at time 1 become more complicated than they are at time 0.

4. Empirical Implementation

4.1 Model Specification

For empirical estimation purposes it is necessary to make some simplifying assumptions. As indicated above, we replace the general stochastic production function (1) by the additive stochastic power production function:

$$Y_t^i(\hat{a}) \equiv \hat{Y}_t^{i\hat{a}} = \hat{e}^{i\hat{a}}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_f} + \theta^{i\hat{a}}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_g} + \sigma(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_h}\varepsilon_t^i. \tag{12}$$

We interpret the output $\hat{Y}_t^{i\hat{a}}$ from the production function as firm (shareholder) wealth at the end of period t. This is made up of two components, the income generated as a consequence of the CEOs effort and ability applied to capital stock $\hat{K}_{t-1}^{i\hat{a}}$ he commands, and the value of the capital stock he commenced with at the beginning of the period, $\hat{K}_{t-1}^{i\hat{a}}$. Naturally, the end of period wealth, $\hat{Y}_t^{i\hat{a}}$, is made up of firm wealth, $\hat{K}_t^{i\hat{a}}$, plus dividends paid to shareholders. Since we cannot observe realized CEO effort we follow the approach of Baker and Hall (2004) and utilize instead the equilibrium level of effort, $\hat{e}^{i\hat{a}}$, for CEO $\hat{a}$ implied by the CEOs contract.

CEO stints with a particular firm are assumed to have a minimum length of two financial years and continue until the CEO resigns, retires or dies. We adopt as our unit of account a particular CEO stint. Recognizing that the length of an individual stint, $n^{\hat{a}}$, varies, the annual firm income, equilibrium CEO effort levels and annual opening capital stock values are averaged over the length of the CEOs stint. The superscript $i\hat{a}$ refer to the mean values over the career of the $\hat{a}$ th CEO in firm i and the subscript t-1 to the beginning of fiscal year opening value. If a particular CEO has (non-overlapping) stints over two or more firms (or separate stints in the same firm with a gap of a minimum of two years) then that CEO will appear more than once in the database.

The analysis is based on two sets of accounts: market and accounting. With respect to our market measure and in keeping with the regression analysis of Gabaix and Landier (2007, Table I), we use the opening market value of total assets, market capitalization plus the value of total debt, as the best size proxy for the capital stock measure that is most associated with CEO pay, rather than income or sales that Gabaix and Landier (2007) show are inferior in their ability to explain CEO pay.

However, when we utilize the alternate measure, namely net accounting income (before interest) as the performance measure we use the opening book value of total assets instead of market capitalization as our size measure.

From the second-period equilibrium pay sensitivity equation (8) the CEO's average equilibrium effort level is

$$\hat{e}_t^{i\hat{a}} = \frac{\hat{\beta}_t^{i\hat{a}}}{\kappa}\left(\hat{K}_{t-1}^{i\hat{a}}\right)^{\gamma_f}, \tag{13}$$

where the superscript $\hat{}$ indicates the mean value of the variable over the CEOs career with the firm. Hence $\hat{\beta}_t^{i\hat{a}}$ is the estimated average pay-performance sensitivity and $\hat{K}_{t-1}^{i\hat{a}}$ the average opening value of assets under management over the CEO's career.

Here we use the averaged sum of the executive's shareholding, restricted stock and the share-equivalent of the executive's option holdings relative to total shareholdings estimated from the Black-Scholes Delta formula (modified to include dividends) to provide the estimated sensitivity for each executive's career. Substituting this expression into the firm's production function (12) we obtain the new expression:

$$\hat{Y}_t^{i\hat{a}} = \frac{\hat{\beta}_t^{i\hat{a}}}{\kappa}\left(\hat{K}_{t-1}^{i\hat{a}}\right)^{2\gamma_f} + \theta^{\hat{a}}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_g} + \sigma(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_h}\varepsilon_t^i, \tag{14}$$

for which we provide direct estimates of the parameters of this non-linear production function below.

Equation (14) is estimated as two separate components. The first component is:

$$\hat{Y}_t^{i\hat{a}} = \frac{\hat{\beta}_t^{i\hat{a}}}{\kappa}\left(\hat{K}_{t-1}^{i\hat{a}}\right)^{2\gamma_f} + \overline{\theta}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_g} + Controls + \xi_t, \tag{15}$$

where $\overline{\theta}$ is the mean (expected) talent factor over all CEO stints in the sample, Controls consist of both two digit Industry Dummies and Year Dummies and the length of experience with the firm prior to the CEO appointment if an internal appointment, and ξ_t is the *iid* error term. For space reasons the control regression parameters are not shown in the tables. Note that the parameters of the dollar volatility term in Eq. (14), σ and γ_h, need to be estimated separately as the term e in dollar volatility $\sigma\left(K_t^{ij}\right)^{\gamma_h} e$, is a standard normal random variable with mean zero. These are estimated via Eq. (16) below.

Our sample consists of 3,238 careers of CEOs in non-financial services firms that have appeared in S&Ps ExecuComp over the period 1992-2005 with no missing observations and a minimum career length of two full financial years. ExecuComp includes every firm over this period that has ever appeared within the top 1500 S&P firms. Moreover, as noted, financial services firms (SIC codes 60 and 61) were eliminated on the grounds that CEO performance characteristics from this sector may not be comparable with the remaining sectors. The Delta value from the Black-Scholes formula is used to compute the share equivalent of option holdings of CEOs, as in Garvey and Swan (2002), to add to regular and restricted stock holdings in computing the average pay-performance sensitivity, $\hat{\beta}_t^{i\hat{a}}$, over the CEOs career with the *i*th firm. Hence a CEO who moves between firms, or who resigns and is rehired two or more years later as CEO with the same firm, will have more than one career recorded.

Two digit SIC code dummies are utilized as industry controls in all the regressions along with year dummy variables. In addition, the CEOs number of years experience with firm prior to becoming the CEO is utilized as a control when ExecuComp records such information. Otherwise, it is assumed that the CEO was hired either externally or with little firm-specific knowledge prior to assuming the role. In the interests of space, the coefficients of control variables are not reported.

The CEO career performance, $\hat{Y}_t^{ij}$, is measured in two ways, based on market and accounting performance respectively:

First, the dollar value of the firm's fiscal year contribution to shareholders is computed as the change in shareholder plus debt-holder wealth (the dollar stock capital gains plus dividends from CRSP plus interest payments on total debt from S&Ps Compustat). Interest payments on total debt are included in the performance measure as these are used in part to fund the total market value of assets which represents the firm's most appropriate size measure from the perspective of the CEOs contribution to managing assets under his control. This firm market-based dollar income measure is then averaged over the period of the CEOs tenure with the firm to obtain the career stint average contribution. This income measure is then added to the mean value of the opening capital stock value over the CEOs stint to obtain the mean stint wealth contribution. All dollar amounts including the value of assets, the firm's total market and accounting income and the CEOs total pay are converted to constant dollars of 2006 using the CPI.

The second income measure is obtained as the fiscal year accounting income on total assets (rate of return on total assets times the book value of total assets) as reported by Compustat. Once again, this dollar performance measure is averaged over the career of each CEO with the firm. Both performance measures are deflated by the estimated average pay-performance sensitivity and then logged to obtain the dependent variable in regression Eq. (5) in the Appendix.

The firm's total value of assets is found as the sum of the fiscal year end market capitalization of the firm from CRSP plus the total value of all debt from Compustat. These values are then averaged over the tenure of each CEO with that firm to obtain the mean value, $K_{t-1}^{i\hat{a}}$. The sample of CEO stints is ranked by the size of the opening total asset value to split the sample into two halves, CEO stints in large and small firms, respectively. The sample utilizing market values is also split into CEO stints with positive and with negative average income values.

All productivity and associated data for CEO careers based on market performance values are summarized in Table 1 and for accounting values, in Table 2. The partial correlation coefficients (in levels) for the entire market-based sample are provided in Table 3.

<< INSERT TABLES 1 to 3 ABOUT HERE>>

The non-linear CEO productivity regression results in real terms are summarized in Table 4 for the two productivity measures and the large and small sub-samples. The coefficients of the control variables are not reported. For the full samples the results show for the market-based measure that CEO effort productivity increases by 55.5 percent for each doubling of firm size (total assets under management) while ability productivity, which can be either positive or negative, increases by a much higher 108.5 percent. The estimated mean CEO ability level for the market measure and full sample is 0.346 which is less the mean of the simulated values of 0.53 found by treating the estimating equation (15) as an identity. All estimates using market returns and market asset values are significant at the 1% level, including the estimated cost (unobservable shadow price) of CEO effort, Kappa, equal to 2.16 overall based on market measures but the price of effort is far lower in small firms at 0.468. As far as we are aware, this is the first time that this variable has been estimated as the methodology of Baker and Hall (2004) does not allow this to be estimated. They assume a value of one.

<< INSERT TABLE 4 ABOUT HERE>>

The scale economy parameter for CEO effort, γ_g, is higher in large firms at 53 percent than in small firms (36 percent) but the remaining parameters are quite similar to that of the full sample. Apart from the Kappa value, all the remaining coefficients are significant at the 1% level. Surprisingly, we find using the market measure that average CEO talent is higher in small firms (0.46 as opposed to 0.37) but this could be due to the technology (parameter) differences between the two types of firms. Hence these findings do not support Rosen's (1982) conjecture that more able managers are systematically employed by larger companies. However, contrary to the Gabaix and Landier (2007) estimates that found negligible differences in ability levels from the CEO in their median company, number 250, in size and number one, we find a remarkable diversity in CEO talent as measured by *ex post* performance. We find that the volatility of ability in large firms to be 9.8 percent, rising to 13 percent in small firms. These findings are consistent with our model in which the CEOs own ex ante ability many be unknown even to himself and where in the marketplace for CEOs it is possible that more capable managers are priced out of the market within the group of large companies.

To obtain the third element in the production function, observations on average dollar volatility of the firm during the careers of each CEO are used to estimate the elasticity of the stochastic production function with respect to volatility:

$$log\left(\text{average career dollar volatility}_{\text{t}}\right) = log\left(\bar{\sigma}\right) + \gamma_h log\left(K_{t-1}^{i\hat{a}}\right), \qquad (16)$$

utilizing the original 3,238 observations and the split samples of large and small firms that form the basis of the Eq. (16) estimates. These regression results are also summarized in Table 4 above. The results indicate that productivity is extremely sensitive to share price volatility with the scale elasticity ranging between 88 percent for the sample of large firms based on market values and as low as 57 percent for small firms based on accounting values.

The final question to be addressed is how total CEO pay responds to both increases in total assets under management and to ability differences. While it is well-established that CEO pay is higher in larger companies, we are not aware of studies showing the responsiveness of pay to differences in ability levels. The address these questions the log of both size and ability are regressed on the log of a comprehensive measure of CEO average total pay over their career with each firm in constant 2006 dollars:

$$\begin{aligned}&\log\left(\text{total pay}_{\text{t}}\right)=\log\left(\text{fixed pay}_{\text{t}}^{\hat{a}}\right)+\rho_{\theta}\theta^{\hat{a}}+\rho\log\left(\beta^{i\hat{a}}\right)+\rho_{K}\log\left(K_{t-1}^{i\hat{a}}\right)\\&+\rho_{yrs\ in\ office}Career\ length_{t}+\rho_{pre\ CEO\ \exp}Exp+Resignation\ Dum_{t}+Conrols+\varepsilon_{t}\end{aligned}, \quad (17)$$

where ρ_{θ} is the elasticity with respect to ability and ρ_{K} is the elasticity of pay with respect to size. A comprehensive measure of total pay is used. It consists of salary plus bonus plus long-term incentive plan plus the value of new options and restricted stock allocated and the change in value of existing option holdings. These values are averaged over each CEOs career. The results are summarized in Table 5.

<< INSERT TABLE 5 ABOUT HERE>>

The impact of the estimated talent for each executive stint on pay is shown in the second row of the Table. These impacts estimates ranging from 1.36 for large firms to 0.396 for small firms based on market productivities. The market talent estimates for the entire sample are 0.67 and thus represent the mixture of the two firm types. Thus large firms rely more on market-based talent measures. For accounting talent measures pay is highly significant at 2.65 for large firms but insignificant for small firms. Market estimates are also provided based on the income productivity of CEO stints split into positive and negative outcomes. These estimates show that talent is significantly rewarded for CEOs with positive ability but impact on pay for those with negative ability is insignificant. Moreover, years in office is significantly rewarded for positive performers and in a negative direction for poor performers.

CEO pay is estimated to increase by between 42 to 53.5 percent for each doubling in total assets, depending on the nature of the productivity measure. These findings are consistent with the elasticity estimates for the scale impacts of effort on productivity.

The elasticity estimates for years of experience with the firm prior to CEO appointment indicate that external appointees are paid more and that this experience is not rewarded. Finally, CEOs with positive performance who die in office suffer a sizeable loss, most likely because the value of options drops. There appears to be no such loss for non-performers.

5. Conclusions

Our modeling shows that when it is sufficiently productive, the large firm expectedly pays higher salary than the small firm. (See Proposition 5). In a neoclassical world, it can be socially optimal that large firms (or firms with better production technologies) hire managers with high abilities. Thus, it is conventionally argued that there should be a positive relationship between pay and firm size because large firms hire high ability managers who deserve high pay. However, in an agency world, a large firm (with better production technology) may not always be willing to hire a high (expected) ability manager partly because most ability rent belongs to the agent in labor market competition and partly because salaries are affected by both the agent expected ability and its volatility. We argue that even when a large firm hires a low-ability manager, the expected pay for the low ability manager can be higher than that for a high-ability manager who is hired by a small firm, if the large firm's productivity and the firm size are sufficiently higher and larger than those of the small firm. We find that, indeed, larger firms have a stronger incentive to hire more able managers as the scale elasticity of ability is positive and significant at 108 percent based on the market performance measure.

We also find that unlike Jensen and Murphy (1990) or Baker and Hall (2004), one may not claim a negative relationship between the sensitivity and the firm size without looking at relative productivities across firms. When we check these estimated productivities we find, indeed, that the sensitivity relationship with firm size is optimally negative in equilibrium.

We analyze managerial career paths which can also affect the contract sensitivity. Since we find that managerial ability contribution increases faster than the total market productivity volatility as the firm size increases, we expect that a manager who previously worked for a large firm will be given a contract with a higher sensitivity than a manager who previously worked for a small firm. We find some limited support for this hypothesis.

References

Baker, G., M. Jensen and K. J. Murphy, 1988, "Compensation and Incentives: Practice vs. Theory," *Journal of Finance* 43 (3), 593-616.

Baker, G. and B. J. Hall, 2004, "CEO Incentives and Firm Size," *Journal of Labor Economics*, 22 (4), 767-798.

Barro, Jason, and Robert Barro, 1990, "Pay, Performance and Turnover of Bank CEOs," *Journal of Labor Economics* 8, 448-481.

Bayard, Kimberly and Kenneth R. Troske, 1999, "Examining the Employer-Size Wage premium in the Manufacturing, Retail Trade, and Service Industries Using Employer-Employee Matched Data," *American Economic Review* 89, 99-103.

Bebchuk, Lucian Arye and Yaniv Grinstein, "Firm Expansion and CEO Pay," Law and Economics, Harvard University Discussion paper No. 533.

Chung, Kim-Sau and Péter Esö, 2007, "Signalling with Career Concerns," Working Paper, School of Economics and Finance, The University of Hong Kong and MEDS Department, Kellogg School, Northwestern University (January).

Conyon, M., 1997, "Corporate Governance and Executive Compensation," *International Journal of Industrial Organization* 15 (4), 493-510.

Conyon, Martin J. and K. J. Murphy, "The Prince and the Pauper? CEO Pay in the United States and United Kingdom," *Economic Journal* 110, 640-671.

Cosh, Andrew, 1975, "The Remuneration of Chief Executives in the United Kingdom," *Economic Journal* 85, 75-94.

Cosh, A. and A. Hughes, 1997, "Executive Remuneration, Executive Dismissal and Institutional Shareholdings," *International Journal of Industrial Organization* 15 (4), 469-492.

Edmans, Alex, Xavier Gabaix and Augustin Landier, 2007, "A Calibratable Model of Optimal CEO Incentives in Market Equilibrium," Working Paper, MIT Sloan School of Management.

Fama, E., 1980, "Agency problems and the theory of the firm," *Journal of Political Economy* 88 (2), 288-307.

Frydman, Carola and Raven E. Saks, 2007, "Executive Compensation: A New View from a Long-term Perspective, 1936-2005," Working Paper, Federal Reserve Board, Washington DC.

Gabaix, Xavier and Augustin Landier, 2007, "Why Has Ceo Pay Increased So Much," forthcoming in the *Quarterly Journal of Economics*.

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

Holmstrom, Bengt, 1999, "Managerial Incentive Problems: A Dynamic Perspective," *Review of Economic Studies* 101, 835-860.

Holmstrom, B. and P. Milgrom, 1987, "Aggregation and Linearity in the Provision of Intertemporal Incentives," *Econometrica* 55, 303-328.

Idson, Todd L., and Walter Y. Oi, 1999, “Workers are More Productive in Large Firms,” *American Economic Review* 89 (2), 104-108.

Izan, H. Y., B. Sidhu, and S. Taylor, 1998, Does CEO pay reflect performance? Some Australian evidence, *Corporate Governance: an International Review* 6, 39–47.

Jensen, M. and K. J. Murphy, 1990, “Performance pay and top-management incentives,” *Journal of Political Economy* 98, 225-264.

Joskow, Paul, Nancy Rose, and Andrea Shepard, 1993, Regulatory Constraints on CEO Compensation,” *Brookings papers on Economic Activity: Microeconomics* MCMXCIII, 1-72.

Kaplan, S., 1997, “Top Executive Incentives in Germany, Japan and the US: A Comparison,” Working Paper, University of Chicago.

Kaplan, Steven N., Mark M. Klebanov and Morten Sorensen, 2007, “Which CEO Characteristics and Abilities Matter?,” Working Paper, University of Chicago.

Kaplan, Steven N. and Joshua Rauh, 2007, “Wall Street and Main Street: What Contributes to the Rise in the Highest Incomes?,” Working Paper, University of Chicago.

Kato, T., 1997, “Chief Executive Compensation and Corporate Groups in Japan: New Evidence from Micro Data,” *International Journal of Industrial Organization* 15 (4), 493-510.

Kato, T. and K. Kubo (2006). "CEO compensation and firm performance in Japan: Evidence from new panel data on individual CEO pay,” *Journal of the Japanese and International Economies*; 20(1): 1

Kostiuk, P. F., 1989, “Firm size and executive compensation,” *The Journal of Human Resources* 25 (1), 90-105.

Lewellen, W. G. and B. Huntsman, 1970, “Managerial pay and corporate performance,” *American Economic Review* 60 (4), 710-720.

Lewis, H.G.,1986, *Union relative wage effects*. Chicago: University of Chicago Press.

McKnight, Phillip J. and Cyril Tomkins, 1999, “Top Executive Pay in the United Kingdom: A Corporate Governance Dilemma,” International Journal of the Economics of Business 6, 223-243.

Merhebi, R., K. Pattenden, P. L. Swan and X. Zhou, 2006, “Australian CEO Remuneration: Pay and Performance,” *Journal of Accounting and Economics*.

Murphy, K. J., 1985, “Corporate performance and managerial remuneration,” *Journal of Accounting and Economics* 7 (2), 11-42.

Roberts, D. R., 1956, “A general theory of executive compensation based on statistically tested propositions,” *Quarterly Journal of Economics* 70 (3), 270-294.

Rose, Nancy, and Andrea Shepard, 1997, “Firm Diversification and CEO Compensation: Managerial Ability or Executive Entrenchment?” RAND Journal of Economics 28, 489-514.

Rosen, S., 1982, “Authority, control and the distribution of earnings”, *The Bell Journal of Economics* 13 (2), 311-323.

Schättler, H. and J. Sung, 1993, “The First-Order Approach to Continuous-Time Principal Agent Problem with Exponential Utility,” *Journal of Economic Theory* 61, 331-371.

Winn, D. N. and J. D. Shoenhair (1988), “Compensation-based (dis)incentives for revenue-maximizing behavior,” *Review of Economics and Statistics*, 70 (1), 154-158.

Zhou, X., 1999, “CEO Pay, Firm Size, and Corporate Performance,” *Canadian Journal of Economics*, 33, 213-251.

Zhou, X. and P. L. Swan, 2003, “Performance Thresholds in Incentive Contracts”, *Journal of Business 76, 665-696.*

APPENDIX

Proofs

Proof of Proposition 1: Note that the sign of the performance sensitivity with respect to firm size:

$$sign\left(\frac{\partial}{\partial K^i}\left(\frac{\kappa\hat{e}}{(K^i)^{\gamma_f}}\right)\right) = -sign\left(\frac{(\gamma_g-\gamma_f)(K^i)^{2(\gamma_g-\gamma_f)-1}\sigma_\theta^2}{\sigma_\theta^2(K^k)^{2(\gamma_g-\gamma_h)}+\sigma^2}+(\gamma_h-\gamma_f)(K^i)^{2(\gamma_h-\gamma_f)-1}\right)$$

$$= -sign\left(\frac{(\gamma_g-\gamma_f)(K^i)^{2(\gamma_g-\gamma_h)}\sigma_\theta^2}{\sigma_\theta^2(K^k)^{2(\gamma_g-\gamma_h)}+\sigma^2}+(\gamma_h-\gamma_f)\right)$$

$$= -sign\left((\gamma_g+\gamma_h-2\gamma_f)-(\gamma_g-\gamma_f)\frac{\sigma^2}{\sigma_\theta^2(K^k)^{2(\gamma_g-\gamma_h)}+\sigma^2}\right).$$

This quantity is $<\ 0$ if $\gamma_h-\gamma_f\ \geq\ \gamma_f-\gamma_g>0$, and $>\ 0$ if $0<\gamma_g-\gamma_f\ \leq\ \gamma_f-\gamma_h$.

The rest of the statement of the proposition is obvious. □

Proof of Proposition 3: By condition (i), we have $\pi^S(\hat{a},W_1^{\hat{a}S})\geq\pi^S(\hat{b},W_1^{\hat{b}L})$. Thus,

$$g(K^S)E[\theta^{\hat{a}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^S)-W_1^{\hat{a}S}\geq g(K^S)E[\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{b}}},K^S)-W_1^{\hat{b}L},$$
$$g(K^S)E[\theta^{\hat{a}}-\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^S)-\Phi(K^{k_{\hat{b}}},K^S)\geq W_1^{\hat{a}S}-W_1^{\hat{b}L}.$$

We also have:

$$\pi^L(\hat{b},W_1^{\hat{b}L})\geq\pi^L(\hat{a},W_1^{\hat{a}S}),$$
$$g(K^L)E[\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{b}}},K^L)-W_1^{\hat{b}L}\geq g(K^L)E[\theta^{\hat{a}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^L)-W_1^{\hat{a}S},$$
$$W_1^{\hat{a}S}-W_1^{\hat{b}L}\geq g(K^L)E[\theta^{\hat{a}}-\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^L)-\Phi(K^{k_{\hat{b}}},K^L),$$

Combining the above inequalities, we have:

$$g(K^S)E[\theta^{\hat{a}}-\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^S)-\Phi(K^{k_{\hat{b}}},K^S)\geq W_1^{\hat{a}S}-W_1^{\hat{b}L}$$
$$\geq g(K^L)E[\theta^{\hat{a}}-\theta^{\hat{b}}\,|\,Y_1]+\Phi(K^{k_{\hat{a}}},K^L)-\Phi(K^{k_{\hat{b}}},K^L),$$

that is,

$$E[\theta^{\hat{a}}-\theta^{\hat{b}}\,|\,Y_1]\leq\frac{1}{g(K^L)-g(K^S)}\int_{k_{\hat{a}}}^{k_{\hat{b}}}\int_{K^S}^{K^L}\Phi_{K^kK^i}(K^k,K^i)dK^idK^k. \qquad \text{(A1)}$$

If $(\hat{a},\hat{b})=(a,b)$, then this inequality implies $E[\theta^a-\theta^b \mid Y_1]\le A(K^S,K^L)$.

If $(\hat{a},\hat{b})=(b,a)$, then the same inequality implies $E[\theta^a-\theta^b \mid Y_1]\ge A(K^S,K^L)$.

On the other hand, by the definition of equilibrium in the executive labor market, reservation certainty equivalent wealth levels $(W_1^{\hat{a}S},W_1^{\hat{a}L})$ are determined as follows:

$\pi^S(\hat{a},W_1^{\hat{a}S})=0,$ and $\pi^S(\hat{b},W_1^{\hat{b}L})=0.$ That is,

$$\begin{aligned}
\pi^S(\hat{a},W_1^{\hat{a}S}) &= g(K^S)E[\theta^{\hat{a}} \mid Y_1^{k_{\hat{a}}}(\hat{a})]+\Phi(K^{k_{\hat{a}}},K^S)-W_1^{\hat{a}S}=0,\\
\pi^S(\hat{b},W_1^{\hat{b}L}) &= g(K^S)E[\theta^{\hat{b}} \mid Y_1^{k_{\hat{b}}}(\hat{b})]+\Phi(K^{k_{\hat{b}}},K^S)-W_1^{\hat{b}L}=0.
\end{aligned}$$

Therefore, the assertion of the proposition follows. □

Proof of Proposition 4: Suppose that at time 0, agent *j* is hired by firm *i*. Then by Proposition 3, we know that agent *j* will move to firm *k* (=*S*,*L*) for the second period with a certainty equivalent wealth of $\tilde{W}_1^{kj}(=g(K^S)E[\theta^j \mid Y_1^k(j)]+\Phi(K^k,K^S))$. Thus, given contract $S_1^i(j)=\alpha^{ij}+\beta^{ij}Y_1^i(j)$, the agent's expected utility at time 0 is:

$$\begin{aligned}
& E\left[-\exp\left\{-r\left(S_1^i(j)-c(\hat{e})+\tilde{W}_1^{kj}\right)\right\}\right]\\
&= E\left[-\exp\left\{-r\left(\alpha^{ij}+\beta^{ij}Y_1^i(j)-c(\hat{e})+g(K^S)E[\theta^j \mid Y_1^i(j)]+\Phi(K^i,K^S)\right)\right\}\right]\\
&= E\left[-\exp\left\{-r\begin{pmatrix}\alpha^{ij}+\Phi(K^i,K^S)-c(\hat{e})+g(K^S)\left(m_{\theta^j}-p^i\overline{e}f(K^i)-p^i m_{\theta^j}g(K^i)\right)\\ +(\beta^{ij}+g(K^S)p^i)Y_1^i(a)\end{pmatrix}\right\}\right]\\
&= -\exp\left\{-r\begin{pmatrix}\alpha^{ij}+\Phi(K^i,K^S)-c(\hat{e})+g(K^S)\left(m_{\theta^j}-p^i\overline{e}f(K^i)-p^i m_{\theta^j}g(K^i)\right)\\ +(\beta^{ij}+g(K^S)p^i)(\hat{e}f(K^i)+m_{\theta^j}g(K^i))\\ -\frac{r}{2}(\beta^i+g(K^S)p^i)^2\left(\sigma_\theta^2 g^2(K^i)+\sigma^2h^2(K^i)\right)\end{pmatrix}\right\}\\
&= -\exp(-rW_0^j) \qquad \text{(A2)}
\end{aligned}$$

Differentiating with respect to effort level *e*, we have first-period pay sensitivity of:

$$\beta^{ij}=\frac{c'(e)}{f^i}-g(K^S)p^i.$$

Note that in equilibrium, $\hat{e}=\overline{e}$, and thus the definition of W_0^j in Eq. (A2) implies Eq. (9).

Thus, first-period expected pay is:

$$E[S^i(Y_1^i)] = W_0^j - \Phi(K^i, K^S) + c(\hat{e}) - g(K^S)m_{\theta^j} + \frac{r}{2}\left(\sigma_\theta^2 g^2(K^i) + \sigma^2 h^2(K^i)\right)\left(\frac{c'(e)}{f^i}\right)^2,$$

and the expected profit to firm *i* for the first period is:

$$\begin{aligned} E[Y_1^i - S_1^i] &= ef(K^i) + m_{\theta^j}\left(g(K^i) + g(K^S)\right) - W_0^j + \Phi(K^i, K^S) \\ &\quad -c(\hat{e}) - \frac{r}{2}\left(\sigma_\theta^2 g^2(K^i) + \sigma^2 h^2(K^i)\right)\left(\frac{c'(e)}{f^i}\right)^2. \end{aligned} \tag{A3}$$

Then the FOC with respect to effort *e* for firm *i* to maximize expected profit implies:

$$\frac{\kappa e}{\left(K^i\right)^{\gamma_f}} = \frac{1}{1 + r\kappa\left(K^i\right)^{-2\gamma_f}\left(\sigma_\theta^2\left(K^i\right)^{2\gamma_g} + \sigma^2\left(K^i\right)^{2\gamma_h}\right)},$$

and the sensitivity of the contract at time 0 becomes as stated in (10).

On the other hand, by Definition 1-(ii), the equilibrium certainty equivalent wealth of agent *j* is $W_0^j = 2m_{\theta^j} g(K^S) + \Phi(K^S, K^S) + \Psi(K^S)$. Thus by substituting this certainty equivalent wealth and the FOC back into Eq.(A3), we have Eq. (11). □

Proof of Proposition 5: If agent *j* (=*a*,*b*) were hired by the small firm, Proposition 4 implies the net profit to the firm would be as follows: $E[Y_1^S - S^S(j)] = m_{\theta^j}\left(g(K^S) + g(K^S)\right) - W_0^j + \Phi(K^S, K^S) + \Psi\left(K^S\right) = 0.$ Since by Definition 1-(ii), the small firm is indifferent between the two agents in equilibrium, we have $E[Y_1^S - S^S(a)] = E[Y_1^S - S^S(b)] = 0$. Thus, $W_0^a - W_0^b = 2(m_{\theta^a} - m_{\theta^b})g(K^S)$. On the other hand, if agents *j*=(*a*,b) were hired by the large firm, the profit to the firm would be as follows:

$$E[Y_1^L - S^L(j)] = m_{\theta^j}\left(g(K^L) + g(K^S)\right) - W_0^j + \Phi(K^L, K^S) + \Psi\left(K^L\right).$$

However, since agent *b* is hired in equilibrium by assumption, Definition 1-(i) implies $(m_{\theta^b} - m_{\theta^a})\left(g(K^L) + g(K^S)\right) + W_0^a - W_0^b \geq 0.$ Since $W_0^a - W_0^b = 2(m_{\theta^a} - m_{\theta^b})g(K^S)$, we have $(m_{\theta^b} - m_{\theta^a})\left(g(K^L) - g(K^S)\right) \geq 0$. Therefore, $m_{\theta^b} \geq m_{\theta^a}$.

For the second statement, note that since $\gamma_g < 2\gamma_f$, $\gamma_h < 2\gamma_f$, and $\gamma_g < \gamma_h$, we have $\frac{\partial}{\partial K^k}\Phi(K^k, K^S) < 0$, and $\frac{\partial}{\partial K^i}\Psi(K^i) > 0$. Thus,

$$E[S_1^L] - E[S_1^S] = (m_{\theta^b} - m_{\theta^a})g(K^S) + \Psi(K^L) - \Psi(K^S) + \Phi(K^S, K^S) - \Phi(K^L, K^S) \geq 0.$$

Therefore, the assertion follows. □

Starting Values for Non-Linear Estimation

In order to be able to estimate the production function in levels starting values of the parameters are required for non-linear estimation. Eq. (14) is rearranged as:

$$\left\{\hat{Y}_t^{i\hat{a}} - \left[\theta^{\hat{a}}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_g} + \hat{\sigma}(\hat{K}_{t-1}^{i\hat{a}})^{\gamma_h}\varepsilon_t^i\right]\right\}\frac{1}{\hat{\beta}_t^{i\hat{a}}} = \frac{1}{\kappa}\left(\hat{K}_{t-1}^{i\hat{a}}\right)^{2\gamma_f}. \tag{4}$$

We obtain initial starting estimates of the effort and ability production elasticities in turn, beginning with the effort elasticity. These starting values are then used in the direct estimation of the non-linear production function. We take advantage of the fact that the ability of the $\hat{a}$ th agent, $\theta^{\hat{a}}$, is a drawing from a random distribution and thus may not be systematically related to capital stock size $\hat{K}_{t-1}^{i\hat{a}}$ and that ε_t^i is a standard-normal random variable with a zero mean. Thus as an approximation we can take the expression in square brackets on the LHS of Eq. (4) to be zero in expectation. On taking logarithms we now obtain a simple estimable equation using ordinary least squares (OLS):

$$\log\left\{\left[\hat{Y}_t^{ia}\right]\frac{1}{\hat{\beta}_t^{ia}}\right\} = \log\left(\frac{1}{\kappa}\right) + 2\gamma_f \log\left(\hat{K}_{t-1}^{i\hat{a}}\right), \tag{5}$$

with the intercept estimate $\hat{\alpha}_0 = \log\left(\frac{1}{\kappa}\right)$, the (common) marginal cost of effort coefficient, $\hat{\kappa} = e^{-\hat{\alpha}_0}$, the estimated effort elasticity with respect to the production function, $\hat{\gamma}_f = \frac{1}{2}\hat{\alpha}_1$, where $\hat{\alpha}_1$ is the slope coefficient. These values are then used as starting values in the non-linear estimation of the regression equation based on the production function, Eq. (15) in the text. The use of the non-linear approach is true to the assumed additive nature of the specified error structure.

The starting values for the ability elasticity and mean ability level are now estimated as follows: Once again setting the error term ε_t^i to its expected value of zero, we have by rearranging Eq. (14):

$$log\left[\hat{Y}_t^i(\hat{a}) - \frac{\hat{\beta}^{i\hat{a}}}{\hat{\kappa}}\left(\hat{K}_{t-1}^{i\hat{a}}\right)^{2\hat{\gamma}_f}\right] = log\left(\overline{\theta}\right) + \gamma_g log(\hat{K}_{t-1}^{i\hat{a}}), \tag{6}$$

where $\overline{\theta}^j$ denotes the mean level of ability.

Table 1: Summary Statistics of CEO Careers, 1992-2005, Based on Market Productivity

Fiscal year annual average values in constant 2006 dollars based on the CPI over the CEOs stint with a single firm. Sources are S&P ExecuComp, S&P Compustat and CRSP. All CEOs excluding those in financial services and with tenure less than two years are included. Market productivity is based on the dollar value of capital gains plus dividends plus the sum of all interest payments on total debt from Compustat plus the opening total market value of assets. Pay-performance sensitivity (Beta) is based on the sensitivity of bonus payments to stock returns, restricted stock grants and the equivalent stock value of option grants relative to shares outstanding, based on the Black-Scholes formula. Total pay consists of the value of salary plus bonus plus restricted stock grants plus the value of option grants as reported by ExecuComp and converted to 2006 prices. The dollar volatility of the firm's stock was computed from CRSP data as the product of the standard deviation of returns and the opening value of market capitalization for each financial year. All values are averaged over the tenure of each CEO with the firm. The CEOs firm experience prior to being appointed CEO is computed from the date the CEO joined the firm until appointed CEO where this is recorded by ExecuComp. Where this information is not reported the CEO is assumed to have been externally recruited. The average age of the CEO over his tenure is computed for the smaller sample of CEOs for which ExecuComp supplies this information. The larger and smaller firm samples are obtained by equally dividing the entire sample of size-ranked CEO stints.

Variable	No.	Mean	Std Dev	Min	Max
Overall Sample					
Avg Mkt Terminal Wlth ($M)	3,238	12,686	40,772	33	957,204
Avg Mkt Val Ttl Asts ($M)	3,238	12,081	39,634	38	954,236
Avg Beta (PPS)	3,238	0.0390	0.0643	0.0000	0.5944
Avg Total Pay ($000)	3,238	4,806	7,386	0	122,874
Avg Dol Volat ($M)	3,238	2,662	8,464	18	191,258
Career Length (Yrs)	3,238	5.2292	2.9124	2	14
Yrs Exp (Pre-CEO)	3,238	6.4722	8.7401	0	48
Avg. CEO Age	2,181	55.4	7.7	30.0	90.5
Sample of Large Firms					
Avg Mkt Terminal Wlth ($M)	1,619	24,290	55,278	1,455	957,204
Avg Mkt Val Ttl Asts ($M)	1,619	23,153	53,824	2,389	954,236
Avg Beta (PPS)	1,619	0.0240	0.0499	0.0000	0.5370
Avg Total Pay ($000)	1,619	7,419	9,549	0	122,874
Avg Dol Volat ($M)	1,619	4,946	11,524	103	191,258
Career Length (Yrs)	1,619	5.4175	2.9853	2	14
Yrs Exp (Pre-CEO)	1,619	8.5559	10.1807	0	42
Avg. CEO Age	1,053	55.7	7.1	34.5	81.0
Sample of Small Firms					
Avg Mkt Terminal Wlth ($M)	1,619	1,081	677	33	4,623
Avg Mkt Val Ttl Asts ($M)	1,619	1,010	602	38	2,387
Avg Beta (PPS)	1,619	0.0540	0.0729	0.0000	0.5944
Avg Total Pay ($000)	1,619	2,193	2,068	0	24,779
Avg Dol Volat ($M)	1,619	379	318	18	2,905
Career Length (Yrs)	1,619	5.0408	2.8259	2	14
Yrs Exp (Pre-CEO)	1,619	4.3885	6.3632	0	48
Avg. CEO Age	1,128	55.2	8.2	30.0	90.5
Sample of Firms with Positive Market Performace					
Avg Mkt Income ($M)	2,382	1,114	3,393	0	68,252
Avg Mkt Val Ttl Asts ($M)	2,382	12,879	41,791	47	954,236
Avg Beta (PPS)	2,382	0.0374	0.0637	0.0000	0.5683
Avg Total Pay ($000)	2,382	4,994	7,828	0	122,874
Avg Dol Volat ($M)	2,382	2,505	7,306	21	140,606
Career Length (Yrs)	2,382	5.5835	3.0213	2	14
Yrs Exp (Pre-CEO)	2,382	6.8505	9.0254	0	48
Avg. CEO Age	1,693	55.7	7.5	34.0	90.5
Sample of Firms with Negative Market Performace					
Avg Mkt Incone ($M)	856	-813	3,262	-42,185	0
Avg Mkt Val Ttl Asts ($M)	856	9,861	32,817	38	427,602
Avg Beta (PPS)	856	0.0435	0.0656	0.0000	0.5944
Avg Total Pay ($000)	856	4,283	5,959	0	45,597
Avg Dol Volat ($M)	856	3,100	11,062	18	191,258
Career Length (Yrs)	856	4.2430	2.3168	2	14
Yrs Exp (Pre-CEO)	856	5.4194	7.8008	0	40
Avg. CEO Age	488	54.5	8.4	30.0	83.0

Table 2: Summary Statistics of CEO Careers, 1992-2005, Based on Accounting Productivity

Fiscal year annual average values in constant 2006 dollars based on the CPI over the CEOs stint with a single firm. Sources are S&P ExecuComp, S&P Compustat and CRSP. All CEOs excluding those in financial services and with tenure less than two years were initially included. Accounting productivity is the product of the accounting rate of return on total assets and the book value of assets from Compustat plus the initial (opening) accounting value of total assets. Pay-performance sensitivity (Beta) is the sum of the sensitivity of bonus payments to stock returns, restricted stock grants and the equivalent stock value of option grants relative to shares outstanding, based on the Black-Scholes formula. Total pay consists of the value of salary plus bonus plus restricted stock grants plus the value of option grants from ExecuComp converted to prices of 2006. The dollar volatility of the firm's stock was computed from CRSP data as the product of the standard deviation of returns and the opening value of market capitalization for each financial year. All values are averaged over the tenure of each CEO with the firm. The CEOs firm experience prior to being appointed CEO is computed from the date the CEO joined the firm until appointed CEO where this is recorded by ExecuComp. Where this information is not reported the CEO is assumed to have been externally recruited. The average age of the CEO over his tenure is computed for the smaller sample of CEOs for which ExecuComp supplies this information. The larger and smaller firm samples are obtained by equally dividing the size-ranked CEO careers.

Variable	No. Obs.	Mean	Std Dev	Min	Max
Overall Sample					
Avg Mkt Terminal Wlth ($M)	3,238	8,222	30,769	-8,818	699,465
Avg Ttl Asts (Bk Val) ($M)	3,238	7,973	30,231	10	681,646
Avg Beta (income share)	3,238	0.0390	0.0643	0	0.5944
Avg Total Pay ($000)	3,238	4,806	7,386	0	122,874
Avg Dol Volatility ($M)	3,238	2,662	8,464	18	191,258
CEO Career Length (Yrs)	3,238	5.2292	2.9124	2	14
Yrs Exp (Pre-CEO)	3,238	6.4722	8.7401	0	48
Avg. CEO Age	2,181	55.4	7.7	30.0	90.5
Sample of Large Firms					
Avg Mkt Terminal Wlth ($M)	1,619	15,843	42,162	-8,818	699,465
Avg Ttl Asts (Bk Val) ($M)	1,619	15,372	41,457	1,409	681,646
Avg Beta (income share)	1,619	0.0225	0.0479	0	0.5370
Avg Total Pay ($000)	1,619	7,100	9,349	0	122,874
Avg Dol Volatility ($M)	1,619	4,747	11,521	71	191,258
CEO Career Length (Yrs)	1,619	5.3891	2.9699	2	14
Yrs Exp (Pre-CEO)	1,619	8.6455	10.2061	0	42
Avg. CEO Age	1,044	56.0	7.0	34.5	82.5
Sample of Small Firms					
Avg Mkt Terminal Wlth ($M)	1,619	600	383	-8	1,625
Avg Ttl Asts (Bk Val) ($M)	1,619	575	361	10	1,401
Avg Beta (income share)	1,619	0.0556	0.0736	0	0.5944
Avg Total Pay ($000)	1,619	2,512	3,347	0	72,297
Avg Dol Volatility ($M)	1,619	578	1,384	18	45,508
CEO Career Length (Yrs)	1,619	5.0692	2.8457	2	14
Yrs Exp (Pre-CEO)	1,619	4.2989	6.2618	0	48
Avg. CEO Age	1,137	54.9	8.3	30.0	90.5

Table 3: Partial Correlation Coefficient Matrix for Entire Sample Utilizing Market Returns, 1992-2005

	Income	Asset	Beta	Pay	Volat	Career	Exper	Age
Avg Mkt Income ($M)	1							
Total Assets ($M)	0.2894	1						
Beta (income share)	-0.0290	-0.1076	1					
Total Pay (ExecuComp)	0.2028	0.3622	-0.0504	1				
Dollar Volatility	0.1068	0.7207	-0.0879	0.4224	1			
CEO Career Length (Yrs)	0.0746	0.0313	0.1610	0.0679	0.0541	1		
Yrs Experience (Pre-CEO)	0.1146	0.1726	-0.1034	0.0035	0.1147	0.0571	1	
CEO Age	0.0091	-0.0068	0.0741	-0.0780	-0.0387	0.0361	0.1160	1

Table 4: CEO Production Function - Non-Linear Regression Equation Estimates Utilizing the Market-Based and Accounting-Based Productivity Samples

Non-linear regression estimates of the production function are given by Eq. (15) in the text. The dependent variable is the average total dollar market income over CEO tenure for each of the three market samples and total asset accounting income for each of the three accounting samples. The three coefficients, Kappa (κ), Gamma_f (γ_f), and Gamma_g (γ_g), are estimated separately for the entire market sample and entire accounting sample and the two subsamples of large and small stocks. The average Theta ($\overline{\theta}$) (CEO ability) factor is estimated for each of the entire samples and for large and small sub-samples. The mean, standard deviation, maximum and minimum values of Theta (θ^{ij}) are also implied by treating the estimated production function as an identity with differing Theta values for each CEO stint.

Coefficient	All		Large		Small	
	Mkt.	Accg.	Mkt.	Accg.	Mkt.	Accg.
Kappa (κ)	2.1615*	1.959*	2.8267*	1.9325*	0.4682*	0.5067*
(t-value)	(11.52)	(10.23)	(7.84)	(7.36)	(3.08)	-8.04
Gamma_f (γ_f)	0.5248*	0.5018*	0.5332*	0.5054*	0.3388*	0.3607*
(t-value)	(117.14)	(93.97)	(80.79)	(68.86)	(10.59)	(27.96)
Gamma_g (γ_g)	1.0853*	1.0793*	1.0793*	1.088*	1.1123*	1.1603*
(t-value)	(6310)	(8994)	(4592)	(5913)	(1571)	(2655)
Est. Ave. Theta (θ)	0.3456*	0.3741	0.3733	0.3349	0.4642*	0.345*
(t-value)	(477)	(689)	(349)	(450)	(225)	(392)
Slope: Predicted Productivity	0.9879*	0.9913*	0.9788*	0.9828*	0.9379*	0.9256*
(t-value)	(497)	(754)	(370)	(535)	(119)	(247)
RSq	0.9871	0.9943	0.9883	0.9944	0.8976	0.9742
Predicted Theta (θ) Mean	0.5328	0.5683	0.5108	0.4798	0.4964	0.3813
Predicted Theta (θ) Std Dev	0.1369	0.0935	0.0979	0.0604	0.1308	0.0687
Predicted Theta (θ) Max	3.8174	0.9400	2.4731	0.7873	3.3162	0.6585
Predicted Theta (θ) Min	0.1053	-0.7181	0.1730	-0.2353	0.0944	-0.6443
	Dollar Volatility Regression					
Gamma_h ($\gamma_{h)}$	0.856*	0.7469*	0.8762*	0.817*	0.7882*	0.5699*
(t val)	(137.1)	(79.1)	(67.47)	(45.68)	(47.2)	(22.35)
Sigma σ	1.0235	3.8036	0.9741	2.3466*	1.5319	10.657*
(t-value)	(0.2632)	(10.19)	(0.1317)	(3.3829)	(3.0814)	(11.39)
RSq	0.8713	0.7	0.7916	0.6566	0.6530	0.38

*Significant at 1%; **Significant at 5%; ***Significant at 10%

Table 5: Determinants of Career Average Career CEO (Flow) Pay Levels (in Logarithms), 1992-2005

Variable	All		Large		Small		Market	
	Mkt	Acc	Mkt	Acc	Mkt	Acc	Positive	Negative
Intercept (Fixed Pay)	3.8527*	4.5836*	3.5265*	3.1470*	4.286*	5.2596*	4.3896*	4.4711*
(t val)	(21.64)	(19.32)	(9.44)	(5.47)	(16.07)	(15.94)	(23.58)	(16.04)
Predicted Ability (θ)	0.6723*	0.4723*	1.3631*	2.649*	0.3964*	-0.1013	0.2004*	0.0318
(t val)	(6.1846)	(2.22)	(5.69)	(4.37)	(2.96)	(0.3014)	(8.15)	(0.4)
Log Beta (PPS)	0.0879*	0.0585*	0.0943*	0.1044*	0.0893*	0.0436**	0.0747*	0.13*
(t val)	(6.55)	(4.025)	(4.685)	(4.99)	(4.92)	(2.07)	(4.79)	(4.73)
Log Total Assets	0.5274*	0.4668*	0.488*	0.5155*	0.5352*	0.4225*	0.479*	0.5178*
(t val)	(41.48)	(30.02)	(19.91)	(14.9)	(19.34)	(13.1)	(33.97)	(22.82)
Log Years in Office	0.0216	0.0276	0.0458	0.0407	-0.0053	0.0048	0.1178*	-0.1261***
(t val)	(0.6996)	(0.8373)	(0.95257)	(0.8314)	(-0.1288)	(0.1027)	(3.03)	(1.906)
Log Yrs Pre-CEO Exp	-0.0609*	-0.0477*	-0.0499**	-0.0498**	-0.0592*	-0.0467**	-0.0727*	-0.0215
(t val)	(4.64)	(3.42)	(2.54)	(2.634)	(3.2)	(2.14)	(4.78)	(0.8141)
Chair-CEO Duality	0.0631***	0.0854**	0.1078**	0.1096**	0.0345	0.0679*	0.0795	0.0386
(t val)	(1.89)	(2.4)	(1.96)	(1.966)	(0.8348)	(1.44274)	(2)	(0.6133)
Departure-Resigned	0.080***	0.081***	0.143**	0.1459**	0.0188	0.006	0.0726	0.0803
(t val)	(1.83)	(1.74)	(1.96)	(1.97)	(0.3488)	(0.0994)	(1.26)	(1.1859)
Departure-Retired	-0.038	-0.0574	-0.0091	-0.0134	-0.1015***	-0.149**	-0.0493	0.0121
(t val)	(1.0124)	(1.4404)	(0.1745)	(0.2629)	(1.8131)	(2.25)	(1.115)	(0.165)
Departure-Deceased	-0.354*	-0.3712*	-0.2525	-0.27	-0.4501*	-0.413**	-0.349**	-0.2447
(t val)	(3.03)	(2.99)	(1.2492)	(1.3855)	(3.22)	(2.56)	(2.53)	(1.07)
No Observations	3,238	3,238	1,619	1,619	1,619	1,619	2,382	856
RMSQ	0.7	0.74	0.7503	0.744	0.6376	0.7308	0.7059	0.659
RSq	0.54	0.48	0.4255	0.4412	0.35	0.2862	0.5255	0.643

*Significant at 1%; **Significant at 5%; ***Significant at 10%