The Impact of Firm Size on Dynamic Incentives and Investment

Chi, Chang Koo and Choi, Kyoung Jin

Norwegian School of Economics, University of Calgary - Haskayne School of Business

20 May 2016

THE IMPACT OF FIRM SIZE ON DYNAMIC INCENTIVES AND INVESTMENT*

Chang-Koo Chi† Kyoung Jin Choi‡

May 20, 2016

Abstract

Recent empirical studies conclude that small firms have higher but more variable growth rates than large firms. To explore the effect of this empirical regularity on moral hazard and investment, we develop a continuous-time agency model with time-varying firm size. Firm size is a diffusion process with two features: the drift is controlled by the agent's effort and the principal's investment decision, and the volatility is proportional to the square root of firm size. We characterize the optimal contract when both parties have CARA utility. The firm improves on production efficiency as it grows, and wages are back-loaded when size is small but front-loaded when it is large. Furthermore, there is under-investment in a small firm but over-investment in a large firm.

JEL CLASSIFICATION: C73, D82, D86, D92

KEYWORDS: Time-Varying Firm Size, Size-Dependence Regularity, Firm Size Effect, Dynamic Moral Hazard, Investment

*We are grateful to Felipe Aguerrevere, Hengjie Ai, Costas Azariadis, Bruno Biais, Alex David, Alex Edmans, Zhiguo He, Hyeng Keun Koo, Christian Krestel, Rodolfo Manuelli, Alfred Lehar, Yuliy Sannikov, Jaeyoung Sung, Noah Williams, John Zhu, and seminar participants in the 2013 Northern Finance Association Meeting and the 2014 Society of Financial Studies Cavalcade for their helpful discussions and suggestions. We are particularly indebted to the editor and three anonymous referees for their valuable advice and comments.

†Aalto University School of Business and HECER, chang-koo.chi@aalto.fi

‡University of Calgary Haskayne School of Business, kjchoi@ucalgary.ca.

1. Introduction

Recent empirical studies conclude that the dynamics of a firm are *negatively* associated with size. It is now well documented that within an industry, small firms grow faster but have a higher volatility of growth rates than large firms. Interpreting the volatility as corporate risk, this empirical pattern—referred to as the *size-dependence regularity* by Cooley and Quadrini (2001)—implies that the degree of risk a corporation faces depends on its size. Then, in a dynamic agency problem in which firm size changes over time, this regularity has important implications for moral hazard and investment, as both are inextricably linked with the characteristic of shocks. In this article, we propose a continuous-time principal-agent model in which firm size follows a diffusion process with a diminishing volatility of the growth rate, which sheds light on the impact of the regularity on dynamic incentives and investment.

There is a growing body of literature, initiated by He (2009), studying how time-varying firm size affects the structure of an optimal contract in a continuous-time framework. However, for the sake of tractability, most of that literature assumes that firm size evolves according to a geometric Brownian motion which, in contrast with the regularity, entails a constant growth rate volatility.[1] Our model features two main departures from the existing models, and they lead to a distinctive firm size process. First, to describe a firm's growth path, we adopt a capital accumulation model in which the principal can increase firm size through investment, and embed it into a dynamic contracting framework. This gives rise to the drift of firm size controlled by the agent's hidden action and the principal's investment decision. Secondly, to incorporate the regularity, we postulate that the volatility of firm size is proportional to the square root of size.[2] The volatility thereby increases with size, but their relationship diminishes as the firm grows. The model provides a simple framework by which we can explore the impact of the regularity on both moral hazard and investment, and delivers qualitatively different predictions about the optimal contract, depending on firm size.

Specifically, the model describes an environment in which a risk-averse principal delegates the management of a firm to a risk-averse agent and offers the agent a long-term contract with full commitment. The contract specifies a flow of compensation for the agent and a flow of dividends for the principal. At each time t, the firm produces output or cash flow with

[1] In He (2009), the agent's hidden action affects the drift of firm size, but not the volatility; the volatility is assumed to be directly proportional to the current firm size.

[2] With the square-root volatility that gives a flavor of the CIR process in Cox *et al.* (1985), we can easily derive sufficient conditions under which the process reaches zero. This is one advantage of working with the square-root volatility rather than other general increasing concave ones. See Online Appendix for the details.

two inputs: (i) the firm's capital stock k_t which also represents the current firm size and (ii) the agent's costly but unobservable effort. The production technology is multiplicative with respect to these two inputs, so the agent's effort has a bigger impact on the firm's profitability in a large firm. To introduce a moral hazard problem, we assume that production is exposed to shocks proportional to $\sqrt{k_t}$, and this is the only source of uncertainty in our model. The realized output can be used for paying compensation and dividends and for investment to expand firm size. The investment plan, defined as the remaining output after payments to both parties, determines the drift of k_t and, moreover, plays a role in transmitting the production shock to the k_t process. As a result, the volatility of k_t is also proportional to $\sqrt{k_t}$.

In this setup, we characterize an optimal long-term contract which maximizes the principal's expected lifetime payoff accruing from dividends, subject to the standard individual rationality and incentive compatibility conditions. To this end, we first utilize the martingale method developed by Sannikov (2008) to derive a stochastic representation for the agent's continuation payoff q_t. As in the previous literature, q_t plays the role of a state variable, and its volatility provides the agent with an incentive for putting forth the necessary effort. Using a recursive definition of the principal's value function, we then formulate the dynamic contract problem into a Hamilton-Jacobi-Bellman equation. However, as our model involves time-varying firm size, the principal's value function inevitably depends on the two state variables, k_t and q_t. Put differently, given current size and promised value to the agent, the principal has to decide how to control the agent's effort and how to expand her own business.

In general, this two-dimensional problem gives rise to partial differential equations which are often difficult to solve even numerically.[3] For the sake of tractability, we assume that both contracting parties have CARA utility á la Holmström and Milgrom (1987). As is well-known, CARA utility allows us to abstract away from the wealth effect on both sides.[4] The absence of wealth effects on the agent's side implies that the agent's promised payoff q_t does not affect his optimal choice of effort. A more important feature of our framework is that on the principal's side, the absence of wealth effects implies that q_t does not influence her investment decision. Taking advantage of these two implications, we can characterize the optimal contract by a system of ordinary differential equations in terms of k_t only. On top of that, we can provide an explicit formula for the evolution of each state variable. In particular, unlike the previous literature in which the contract is characterized by a function of the agent's continuation value

[3] To circumvent such difficulty, most literature (e.g., He (2009), Biais *et al.* (2010), DeMarzo *et al.* (2012)) exploits the scale-invariance principle that stems from (i) the homogeneity of degree 1 of geometric Brownian motions and (ii) the risk-neutrality of contracting parties.

[4] If the principal is risk-neutral, the $\sqrt{k_t}$ volatility has no meaningful implications for dynamic investment.

per capita (see footnote 3), using the formula for q_t, we can state every contractual policy in terms of firm size only. This enables us to address the question of how a change in firm size affects the optimal contract, which we will refer to as "the firm size effect" hereafter.

We start with the optimal effort policy. The agent's CARA utility leads to the optimal level of effort which is determined by firm size only. This facilitates comparison with the first-best level, so that we can analyze how the degree of distortion arising from moral hazard varies with firm size. As in a standard agency model, the optimal effort is below the first-best level for every firm size. However, in our model, this distortion dwindles away as the firm grows. In Section 4, we prove that the optimal effort converges to the efficient level, and under a very mild condition, the large firm retrieves production efficiency in the sense that there is no wedge in the marginal product of capital between the two environments. To understand this result, note that the agent's effort affects the expected flow of output, which is proportional to k_t because of the multiplicative production technology. On the other hand, the production shock is proportional to $\sqrt{k_t}$. Therefore, in a large firm, effort has a larger effect on the level of output relative to the shock; put simply, the signal-to-noise ratio increases with firm size. This increased ratio enhances informativeness of the realized output about the agent's hidden effort, thereby reducing the cost of incentive provision and also contributing to a negative relationship between pay-performance sensitivity (Jensen and Murphy (1990)) and firm size.

We next investigate the firm size effect on the instantaneous payment scheme (or the flow of compensations) and the agent's continuation value. Unlike the effort policy, the payment scheme naturally depends on both state variables in that the principal should pay for the promised value q_t through the payment. Thus, a change in k_t has a direct effect and an indirect one on the scheme via q_t. To ascertain the exact effect of firm size, therefore, we use the explicit formula of q_t for its relationship k_t, and reformulate the scheme in terms of firm size only. It turns out that q_t keeps track of all histories of the firm's growth, i.e., $\{k_s,\ s \in [0,t]\}$, and thus the reformulated scheme depends not only on the firm's current size but also on its past growth path.[5] This fact creates a link to the previous literature, which studies a dynamic agency problem between parties with CARA utility but assumes time-invariant firm size. The payment scheme in our model aggregates information about the firm's past growth, whereas the lump-sum payment scheme in Holmström and Milgrom (1987) or Schättler and Sung (1993) aggregates information about the agent's past performance.

[5]Recall that in a dynamic moral hazard model with time-invariant firm size, the continuation value contains all records of the agent's past performances. Refer to Spear and Srivastava (1987) for a discrete-time setting and Sannikov (2008) for a continuous-time setting.

In order to compare with the lump-sum payment scheme and highlight the role of time-varying firm size, in Section 4, we decompose our payment scheme into six components in a similar fashion as Holmström and Milgrom (1987). In addition to the standard four components, the scheme comprises two distinctive components which result in interesting wage dynamics. The first one accounts for the adjustment of compensations due to the firm's investment motive. The agent in a small firm is not fully compensated for the cost of effort, and the spare amount of money is spent in investment to expand firm size. Instead of deferring the payment, the principal promises to pay more later by increasing the drift of the continuation payoff. The second component captures exactly the payment from a change in q_t over time. Therefore, when firm size is small, the continuation payoff has a upward drift, implying that wages are back-loaded. When firm size is large, however, the opposite happens: The principal starts paying for her liability by rewarding the agent more than the cost of effort, thereby lowering the drift of q_t. The resulting downward drift of q_t corresponds to front-loaded wages.

Lastly, we consider the optimal investment plan in Section 5. After paying compensation to the agent from output, the principal faces a decision problem of distributing the remaining output into dividend payment for her current interest and investment for future production. In this problem, the continuation value affects the level of the remaining output, but not the return on investment. Accordingly, thanks to the principal's CARA utility, the optimal investment plan is determined by k_t only. Relying on this property, we examine the firm size effect on investment distortions.

When firm size is small, the model predicts under-investment. This is consistent with what is unambiguously predicted by a broad class of agency models, irrespective of firm size. In contrast, our model surprisingly delivers the opposite prediction when firm size is large: A large firm is prone to over-investment. To understand the intuition behind this result, recall that the agency problem leads to a loss in the marginal product of capital in the optimal contract, which is the main driving force of under-investment. But, as discussed earlier, such distortions do not occur in a large firm due to the increased signal-to-noise ratio. Hence all that matters to the risk-averse principal is the amount of risks generated by investment she has to bear, because investment increases the volatility of future production.[6] However, the optimal contract trades off some benefits from risk-sharing for incentive provision, so the principal would be exposed to a smaller amount of risk than the Pareto-efficient one. This reduces the implicit cost of investment, thereby leading to over-investment in the optimal contract. We also discuss the

[6]Thus, if the principal is assumed to be risk-neutral, then there is no investment distortion in a large firm. This is the case in DeMarzo *et al.* (2012) in which the degree of under-investment dwindles away as the state variable, q_t/k_t, rises.

implications of over-investment for marginal Tobin's Q and dividend smoothing.

Literature Review

There has been rapidly growing literature on dynamic contracting such as Biais *et al.* (2010), DeMarzo *et al.* (2012), He (2009, 2011) and other references therein.[7] One of our contributions to the literature is that we investigate the impact of the size-dependence regularity on incentive provisions and investment. The early literature on dynamic contracting employed the arithmetic Brownian motion setting (e.g., DeMarzo and Sannikov (2006) and Sannikov (2008)), in which firm size is assumed to be time-invariant. Afterwards, a strand of literature such as He (2009, 2011), and DeMarzo *et al.* (2012) studied the Geometric Brownian motion (GBM) type model.[8] However, in these models, the volatility of the firm size process linearly increases with size, which is not consistent with the empirical regularity.

The most important feature distinguishing our study from the literature is that we model the diminishing growth rate volatility and explicitly characterize the incentive scheme and investment in terms of firm size only. This characterization has several advantages. First, we can directly investigate how the incentive scheme and investment change as firm size expands or declines. Second, several implications resulting from our characterization are rather easily testable in the aspect that it is unnecessary to come up with an empirical proxy related to the agent's continuation value. In most of the articles that involve time-varying firm size referred to above, the principal's value function becomes homogenous in size. This size-homogeneity enables the agent's continuation value per unit of capital to be a sufficient statistic for characterizing the optimal contract, so it helps to obtain tractable solutions. However, the interesting properties of incentives and investment changing as firm size evolves have largely been simplified. For example, DeMarzo *et al.* (2012) studied dynamic contracting linked with Q theory of investment.[9] Biais *et al.* (2010) study a dynamic moral hazard model with large and infrequent risks. Similar to He (2009), both DeMarzo *et al.* (2012) and Biais *et al.* (2010) investigate the incentive provisions through the size-adjusted continuation value process. However, it is quite challenging in their models to directly extract the firm size effect that is tightly blended with the whole past history of the continuation value (per capita) process. In contrast, the solu-

[7] See also Ai and Li (2015), DeMarzo and Sannikov (2006), and DeMarzo and Fishman (2007).

[8] To our best knowledge, there is one exception He (2011), which permitted a general class of firm dynamics. However, its main focus (the impact of agency problems on firm value and capital structure) is investigated under the GBM model.

[9] In DeMarzo *et al.* (2012), the continuation value (per capital) can be interpreted as a measure of the firm's liquid reserves or financial slack. Although they provided rich implications based on financial slack, past profitability and investment, they are silent about the firm size effect on dynamic incentives and investment.

tion in our model does not exhibit such homogeneity because of the square-root volatility, but we are able to explicitly specify the payment scheme and investment decision in terms of firm size only. We will further discuss the differences between our study and the aforementioned dynamic contracting literature in the main body of the article.

The related literature on dynamic contracting with firm dynamics also includes Albuquerque and Hopenhayn (2004), Clementi and Hopenhayn (2006), and Clementi *et al.* (2010). Albuquerque and Hopenhayn (2004) and Clementi and Hopenhayn (2006) analyze the optimal *debt contract*, by which they account for the size-dependence regularity. Our focus is to investigate the contracting problem between the manger and the shareholders and how the incentive changes under the size-dependence regularity, not to generate such regularity. The focus of Clementi *et al.* (2010) is on dynamics as a firm gets older, and more precisely, the decrease of firm size for old firms, whereas our focus is on the cross-sectional aspect.

There is also a significant body of literature on over- or under-investment issues. Here instead of surveying all those articles, we shall introduce the models with dynamic features in order to narrow the scope. According to Stein (2003), there are two broad categories of literature with respect to the investment issue: one with models of agency conflicts, and the other with models of costly external finance. The former generally predicts over-investment and the latter generally predicts under-investment. For instance, Dow *et al.* (2005) and Albuquerque and Wang (2008) predict over-investment. Dow *et al.* (2005) is based on the free cash flow theory of Jensen (1986). Albuquerque and Wang (2008) consider the agency conflict between the controlling shareholder and outside investors. In both models, investment decision makers such as empire builders or controlling shareholders have incentives to over-invest. The inefficient investment in our model is generated by the deviation from optimal risk sharing for incentive provision, not by the imperfect protection of the shareholders.

On the other hand, the usual dynamic contracting theory referred to above often predicts underinvestment.[10] Albuquerque and Wang (2008) point out that over-investment is likely to be the dominant issue for large firms around the world, whereas the underinvestment implied by these contracting models is potentially more important for small firms. To our knowledge, our model is the first that has both under- or over-investment features depending on firm size. We hope that this article sheds light on the integration of two separate views of the investment decisions of small and large firms.

[10] Probably one exception is Gryglewicz and Hartman-Glaser (2015). In a real option framework, they show that severe moral hazard can lead to an early exercise of real options, which can be interpreted as over-investment in their case.

The remainder of the article is organized as follows. In Section 2 we present the general theme of our model and formulate the optimal contract problem. In Section 3, we characterize the optimal contract and provide an explicit formula for the evolution of each state variable when both contracting parties have CARA preferences and the firm's production technology is multiplicative. In Section 4 we analyze the optimal contract in details with an emphasis on the firm size effect on the effort policy, the incentive scheme, wage dynamics, and pay-performance sensitivity. In Section 5 we analyze the investment plan. Concluding remarks are gathered in Section 6. The first-best contract is characterized in Appendix A, and all omitted proofs are relegated to Appendix B. Online Appendix provides some technical detail on the boundary behavior of the firm size process.

2. The Model

We consider a continuous-time agency model in which a principal (shareholders) delegates the management of a firm to an agent (executives). During any time interval $[t, t + dt)$, the firm produces output or cash flow with two inputs, the agent's effort e_t and the firm's capital stock k_t. Throughout the article, the capital stock k_t will be used as a unique metric to judge firm size or the firm's market value at time t. Also, the firm's production is subject to risks whose volatility is dependent on k_t. Specifically, the cumulative output Y_t up to time t evolves according to

$$dY_t = f(k_t, e_t)dt + \sigma\sqrt{k_t}\, dW_t. \tag{2.1}$$

Here, the drift $f(k_t, e_t)$ represents the firm's production technology, W_t is a Brownian motion in standard probability space $(\Omega, \mathcal{F}, P)$, and the volatility term $\sigma\sqrt{k_t}\; dW_t$ indicates the *size-dependent* production shock. The set of feasible effort levels, denoted $\mathcal{E}$, is a compact set of progressively measurable processes with respect to $\mathcal{F}_t$.

The realized output is publicly observable and verifiable, but the agent's choice of effort $e_t \in \mathcal{E}$ is unobservable to the principal due to the production shock. To provide the agent with an incentive to work, the principal offers a contract with commitment at time $t = 0$. A contract explicitly specifies a flow of compensation c_t to the agent and a flow of dividends d_t to the principal during any time interval $[t, t + dt)$. The flow of compensation yields the agent a flow of utility $u(c_t, e_t)$ when the agent chooses e_t, and the flow of dividends yields the principal

$v(d_t)$. c_t is the only income source for the agent and he cannot save or borrow money.[11]

On top of the payoff, a contract determines the firm's growth path, which describes the evolution of k_t over time. To formulate this path, we adopt a simple capital accumulation model in which the firm's capital stock accumulates by investment $d\mathcal{I}_t$, where $\mathcal{I}_t$ denotes the cumulative investment up to time t with $\mathcal{I}_0 = 0$, but depreciates at a rate of $\delta \in [0, 1)$. That is, k_t evolves according to $dk_t = d\mathcal{I}_t - \delta k_t dt$. Also, as is standard in the capital accumulation model (e.g., Greenwood *et al.* (1997)), instantaneous investment $d\mathcal{I}_t$ is determined by the remaining output after paying pecuniary compensations to both parties.[12] We thereby provide a unified framework in which the agent's incentive scheme and the corporate growth path are simultaneously determined by a simple contract.

Combining with the output process laid in (2.1), we obtain the following stochastic differential equation (SDE) of k_t:

$$dk_t = \underbrace{dY_t - (c_t + d_t)dt}_{=d\mathcal{I}_t} - \delta k_t dt = \Big[f(k_t, e_t) - c_t - d_t - \delta k_t\Big]dt + \sigma\sqrt{k_t}\, dW_t. \tag{2.2}$$

The production shock is therefore transmitted into the k_t process through investment or the remaining output after compensation, and it constitutes the volatility of k_t. Attentive readers may be concerned about the existence of a solution to the SDE (2.2), because the volatility $\sigma\sqrt{k_t}$ is not Lipschitz-continuous. However, the classic result of Yamada and Watanabe (1971) establishes the existence and uniqueness of a solution to a large class of stochastic differential equations, and the class subsumes (2.2) as a special case. The discrete time analogue of our model can be written by

$$k_{t+1} = (1-\delta)k_t + i_t \quad \text{and} \quad i_t = f(k_t, e_t) - c_t - d_t + \sigma\sqrt{k_t}\,\epsilon_t,$$

where i_t is the amount of investment at time t and the random noise ϵ_t is i.i.d. standard normal.

The reason why we work with the square-root process is two-fold. First, the process does not allow k_t to go below zero, which is analogous to a geometric Brownian motion employed by He (2009) for a firm-size process. k_t is bounded below by zero as its volatility is otherwise undefined. For a diffusion process with constant volatility like the arithmetic Brownian motion, there can be substantial drops in k_t and thus the process is unbounded below. Accordingly, the firm is at risk of default during any time interval, which is apparently unappealing.

[11] For this reason, we sometimes call c_t the flow of consumption depending on the context. In Section 6, we briefly discuss how the results would change if we allow the agent to privately save money.

[12] An implicit assumption we make here is that the capital price is normalized to one.

Secondly and more importantly, the square-root process reflects two empirical patterns on the dynamics of a firm. It is well-documented by a large literature, beginning with Hymer and Pashigian (1962), that (i) an aggregate underlying shock induces larger swings in cash flows of large firms than in those of small firms, but (ii) large firms have a lower standard deviation of the growth rate compared to small firms (see also Evans (1987), Hall (1987), Cooley and Quadrini (2001), and Bottazzi *et al.* (2011)). In fact, the SDE laid out in (2.2) captures this size-dependence regularity; the volatility of k_t itself increases with firm size, but the growth rate of k_t

$$\frac{dk_t}{k_t} = \left[\frac{f(k_t, e_t) - c_t - d_t - \delta k_t}{k_t}\right] dt + \frac{\sigma}{\sqrt{k_t}}\, dW_t$$

has a decreasing volatility with size.[13]

In this circumstance, we formally define a long-term contract as a history-dependent triplet $(c_t, d_t, e_t)_{t\in[0,\infty)}$ or simply (c, d, e) by suppressing the time subscripts. The consumption and dividend plans are the explicit part of a contract, whereas the agent's recommended action is the implicit part. We say that a contract is feasible if all contractual terms (c_t, d_t, e_t) at each time t are contingent on the completion of the σ-algebra generated by all possible histories of capital $\{k_s\}_{s\le t}$. We will denote by $\mathcal{S}$ the set of such feasible contracts.

We assume that both parties discount the future payoff at a common rate $\beta \in (0, 1)$. The expected lifetime utility from a contract can be written as

$$V_0(c, d, e) \equiv \mathbb{E}\left[\int_0^\infty e^{-\beta t} v(d_t)\, dt\right] \quad \text{and} \quad U_0(c, d, e) \equiv \mathbb{E}\left[\int_0^\infty e^{-\beta t} u(c_t, e_t)\, dt\right]$$

for the principal and the agent, respectively.

Given the firm-size process (2.2), the principal's problem is then to offer a feasible contract that maximizes her expected lifetime utility, satisfying the two standard constraints: (i) Individual Rationality (IR), that the contract promises the agent a higher expected utility than his reservation utility q_0; and (ii) Incentive Compatibility (IC), that the contract implements the recommended effort process $e = \{e_t\}_{t\in[0,\infty)}$ so the agent can maximize his expected utility by

[13] The standard deviation of the growth rate of a firm, $\sigma(dk/k)$, is known to satisfy the power law, $\sigma(dk/k) \sim e^{bk}$ for some coefficient b. In various industries, the values of b have been estimated as being negative. The negative association is also in line with the macroeconomic literature that studies a relationship between the mean growth rate and its volatility at the country level. Refer to Section 3.2 of Jones and Manuelli (2005) for a survey on this strand of the literature.

following the instruction. Formally, the problem is stated as follows:

$$\max_{(c,d,e)\in\mathcal{S}} \quad V_0(c,d,e)$$

subject to (2.2),

(IR) $\quad U_0(c,d,e) \geq q_0$, and

(IC) $\quad e \in \underset{e'\in\mathcal{E}}{\operatorname{argmax}}\ U_0(c,d,e')$.

Note that the above formulation of (IC) implies sequential incentive compatibility; the agent is willing to follow the instruction at any time t and irrespective of what history occurred up to t. Also, the condition (IR) implicitly assumes that the agent can commit himself to participation in the contract.[14]

Incentive Compatibility

In this subsection we employ the martingale method developed by Sannikov (2008) to characterize the (IC) condition in terms of the agent's continuation value, which is now a central tool in the dynamic contract literature.

Given a long-term contract (c,d,e) and a history $\mathcal{F}_t$ up to time t, we define the continuation value q_t—the agent's expected future payoff promised by the contract—as

$$q_t(c,d,e) \equiv \mathbb{E}^e \left[\int_t^\infty e^{-\beta(s-t)} u(c_s,e_s)\, ds \,\middle|\, \mathcal{F}_t \right],$$

where $\mathbb{E}^e$ indicates the expectation with respect to the probability measure P^e induced by the agent's choice of effort. Using q_t, we can write the agent's expected lifetime utility evaluated at time t as

$$U_t \equiv \mathbb{E}^e \left[\int_0^\infty e^{-\beta s} u(c_s,e_s) ds \,\middle|\, \mathcal{F}_t \right] = \int_0^t e^{-\beta s} u(c_s,e_s) ds + e^{-\beta t} q_t.$$

Key to the martingale method is the fact that the process U_t becomes a P^e-martingale, i.e., $\mathbb{E}^e[U_T|\mathcal{F}_t] = U_t$ for every $0 \leq t \leq T$. Thus, by the martingale representation theorem, U_t can be represented as an Itô's integral, $U_t = U_0 + \int_0^t e^{-\beta s}\, \Gamma_s\, \sigma \sqrt{k_s}\, dW_s^e$ for a progressively measurable

[14] If the agent could quit at any time of the contract, then we need additional participation constraints that require the agent be assured of his reservation utility at each history, that is, $q_t(c,d,e) \geq q_0$ for all $\mathcal{F}_t$ and $t \geq 0$. We briefly discuss in Section 6 how the agent's inability to commit himself to participating affects our results.

process Γ_t and a Brownian motion W_t^e under P^e. This provides the following dynamics of q_t and characterization of (IC) as well.

PROPOSITION 1. *Given a feasible contract $(c, d, e) \in \mathcal{S}$, there exists a progressively measurable process $\{\Gamma_t\}_{t\in[0,\infty)}$ such that the agent's continuation value q_t evolves according to*

$$dq_t = \left(\beta q_t - u(c_t, e_t)\right)dt + \Gamma_t \underbrace{\left(dY_t - f(k_t, e_t)dt\right)}_{=\sigma\sqrt{k_t}\,dW_t^e} \tag{2.3}$$

with $\mathbb{E}^e[\int_0^t \Gamma_s^2 ds] < \infty$ for all $t \in [0, \infty)$. The contract satisfies the (IC) condition if and only if

$$e_t \in \underset{e'\in\mathcal{E}}{\operatorname{argmax}}\; u(c_t, e') + \Gamma_t f(k_t, e') \tag{2.4}$$

for all $t \in [0, \infty)$ and P^e-almost surely.

PROOF OF PROPOSITION 1: The evolution of q_t is derived from the two expressions of U_t above. Differentiating them with respect to t, equating the two derived equations of dU_t, and then solving for dq_t gives (2.3). The proof of characterization of (IC) is relegated to Appendix B. □

The drift part of q_t follows from the promise-keeping condition; $u(c_t, e_t)dt + dq_t$, the total flow of utility during $[t, t + dt)$, must increase at a rate of βq_t over time. The volatility part Γ_t of q_t, on the other hand, measures sensitivity of the process in response to a change in output dY_t, and thus it provides the agent with an incentive to work and plays a crucial role in characterization of the (IC) condition. Furthermore, the size-dependent production technology in (2.4) suggests that k_t also affects the agent's choice of optimal effort and, in turn, the consumption plan. Note also that the objective function in (2.4) exhibits complementarity between the choice variable e' and parameter Γ, implying that a higher volatility leads to a higher level of effort.

The HJB Equation

We now use the results in Proposition 1 to restate the optimal contract problem into a recursive form. As in the other dynamic moral hazard literature stemming from Spear and Srivastava (1987), the agent's continuation value serves as a state variable that determines the contractual terms (c, d, e) and controls the agent's hidden action. In addition, because our model involves time-varying firm size, the principal's value function inevitably depends on k_t as well.

Indeed, the two variables (k_t, q_t) provide a Markovian structure with our model in the sense that the two variables keep a record of full histories up to t, so the principal can design the forward contractual terms on the basis of k_t and q_t only. For this reason, we denote by $J(k_t, q_t)$ the principal's continuation value function, that is, her expected maximum payoff from time t on given a state (k_t, q_t). Let k_0 denote an initial firm size and q_0 the agent's reservation utility. We then write the optimal contract problem as follows.

$$J(k_0, q_0) \equiv \max_{(c,d,e)\in\mathcal{S}} V_0(c, d, e)$$

subject to the two SDEs:

$$\text{(2.2)} \quad dk_t = [f(k_t, e_t) - c_t - d_t - \delta k_t]dt + \sigma\sqrt{k_t}\, dW_t$$

$$\text{(2.3)} \quad dq_t = [\beta q_t - u(c_t, e_t)]dt + \Gamma_t\, \sigma\sqrt{k_t}\, dW_t \quad \text{with } \Gamma_t \text{ satisfying (2.4).}$$

Using the recursive structure, we can reformulate the above problem into the following Hamilton-Jacobi-Bellman (HJB, hereafter) equation:

$$\text{(HJB)} \quad \begin{aligned} \beta J(k, q) = \max_{(c,d,e)\in\mathcal{S}} \Bigg\{ & v(d) + J_k\Big[f(k, e) - c - d - \delta k\Big] + J_q\Big[\beta q - u(c, e)\Big] \\ & + \frac{1}{2}\Big[J_{kk} + 2J_{kq}\Gamma + J_{qq}\Gamma^2\Big]\,\sigma^2 k\Bigg\}, \end{aligned}$$

where Γ is the volatility of q that satisfies the incentive compatibility condition (2.4) in Proposition 1, and the (double) subscripts of J denote its (second-order) partial derivatives. Intuitively, the principal's expected flow of value $\beta J(k, q)$ on the left side must equal the sum of the instantaneous flow of utility from dividends and the expected change in her continuation value due to the drift and volatility of each state variable.

3. The Optimal Contract

In this section, we characterize an optimal contract. We first specify the contracting parties' utility functions and the firm's production technology. We then conjecture a solution to the HJB equation, derive Euler equations that an optimal contract has to satisfy, and verify that a solution to the equations is indeed optimal.

CARA Preferences and Multiplicative Technology

For the sake of tractability, the general formulation presented in the previous section is specialized as follows. First, we consider a simple multiplicative production function that entails a marginal product of the agent's effort increasing with firm size:

$$f(k_t, e_t) = (k_t + h)e_t,$$

where the parameter $h > 0$ represents the agent's working skills or human capital. In our model, it serves to set the lower bound for the marginal product of the agent's effort. We assume that h is constant over time; there is no learning effect through experience.

Second, we assume the constant absolute risk aversion (CARA) preferences for both parties in the same spirit of Holmström and Milgrom (1987). As is well-known, CARA utility abstracts away from the income effect and thus greatly simplifies our algebra work. Specifically, each party's utility function takes a form of

$$v(d) = -\frac{1}{R}\exp(-Rd) \quad \text{and} \quad u(c,e) = -\frac{1}{r}\exp\left(-r\left(c - \frac{(k+h)e^2}{2a}\right)\right).$$

Here, R and r indicate the constant risk aversion coefficient for the principal and for the agent, respectively. The agent's monetary cost from exerting effort is assumed to be quadratic with respect to e, and the constant $a > 0$ in the denominator determines the optimal level of effort with full information, as is illustrated in Appendix A. This facilitates comparison with the second-best effort policy in the next section. Also, the cost function scales with firm size, reflecting that the agent incurs a higher opportunity cost for management of a large firm.

In order to make our problem interesting, we shall impose the following condition on the parameters defined above:

Assumption 1 (Feller Condition I).

$$\frac{a}{2} > \delta + \beta\left[1 + \frac{Rr\sigma^2}{2(R+r)}\right].$$

This is a version of the Feller condition tailored to our firm-size process with the $\sqrt{k_t}$ volatility, which plays a crucial role in analyzing the process in the first-best contract. To be specific, the condition ensures the drift of k_t to be positive for all $k_t > 0$, so that it prevents the process from reaching its boundary zero almost surely (or the firm from being liquidated; see Section 3). As is illustrated in Online Appendix, the condition is somewhat weaker than the one for

k_t not to reach zero in a finite time. Throughout the remainder of this article we will maintain Assumption 1, although we do not explicitly mention it.

Ordinary Differential Equations

In this subsection, from the HJB equation, we derive Euler equations that characterize the optimal contract. Although the HJB equation is a two-dimensional partial differential equation (PDE), it turns out that the resulting Euler equations can be simplified into a pair of more tractable ordinary differential equations (ODEs) under the CARA utility environment.

We first describe the possibility of terminating a contract or liquidating a firm in line with Sannikov (2008), which is necessary for getting a boundary condition of the ODEs.[15] For an illustration, we assume that at the time of signing a contract, the two parties agree to liquidate the firm when k_t reaches zero. When the firm goes into liquidation, say at $\tau \equiv \inf\{t > 0 | \, k_t = 0\}$, the principal promises the agent a constant flow of severance pay c from τ on, but the agent is allowed to choose zero effort. Hence the agent's flow of utility is $u(c, 0) = -\exp(-rc)/r$. The amount of c the agent receives is determined by his continuation value at the time of liquidation, q_τ. More precisely, the agent receives $c_t = c$ for all $t \geq \tau$ as much as his expected payoff from the flow of utility $u(c, 0)$ from τ on equals the promised payoff by the original contract:

$$\mathbb{E}\left[\int_\tau^\infty e^{-\beta(t-\tau)} u(c, 0)\, dt\right] = q_\tau \quad \text{or} \quad c = \ln(-q_\tau r\beta)^{-\frac{1}{r}}.$$

On the other hand, from the time of liquidation onward, the principal's flow of utility becomes $v(-c) = -\exp(Rc)/R$. Similar to the agent's case, the expected payoff from this flow must match with the principal's continuation value at the liquidation state, that is, $J(0, q_\tau)$. Consequently, we have

$$J(0, q_\tau) = \mathbb{E}\left[\int_\tau^\infty e^{-\beta(t-\tau)} v(-c) dt\right] = -\frac{(-q_\tau r\beta)^{-\lambda}}{R\beta}, \tag{3.1}$$

where $\lambda \equiv R/r$ is the ratio of the risk aversion coefficients. As is revealed shortly, this *value-matching* condition (3.1) translates into a boundary condition.

In light of the principal's CARA utility function and the condition (3.1), we conjecture that the value function takes a form of $J(k, q) = -(-q)^{-\lambda} \exp(-\theta(k))$ for some C^2-function

[15] In Sannikov (2008), there is another boundary condition for the principal's value function, which results from the agent's income effects. When the continuation value is large enough, it is optimal for the principal to retire the agent on account of costly incentives. This type of boundary condition is not necessary in the absence of income effects.

$\theta(k)$. To ensure that the conjectured function is strictly concave in the two variables, we restrict our attention to the set of possible functions $\theta(k)$ satisfying that (i) the function $\theta(k)$ is twice continuously differentiable, (ii) its derivative $\theta'(k)$ is strictly positive and bounded, and (iii) $(\theta'(k))^2 > (\lambda+1)\theta''(k)$ holds for all $k \in [0,\infty)$. Also, it follows from (3.1) that $\theta(0) = \ln\left[R\beta(r\beta)^{\lambda}\right]$. We will later verify that $J(k,q)$ satisfying the above conditions represents the principal's maximum possible value.

We are now ready to derive the desired Euler equations from the HJB equation and characterize the optimal contract.

PROPOSITION 2 (Optimal Contract). *Let $(e^*(k), \theta(k))$ denote a C^1-solution to the system of first-order ordinary differential equations,*

$$e'(k) = F(\theta(k), e(k), k) \quad \textit{and} \quad \theta'(k) = G(e(k), k),$$

satisfying the boundary condition $(e^*(0), \theta(0)) = \left(a, \ln\left[R\beta(r\beta)^{\lambda}\right]\right)$. The exact functional forms of F and G are given in the proof. Define a function ψ^* as

$$\psi^*(k) \equiv 1 + \frac{r(k+h)}{a} e^*(k)(e^*(k)-a),$$

and write the flow of consumption and dividend as

$$c^*(k,q) = \frac{(k+h)e^*(k)^2}{2a} - \frac{1}{r}\ln\left[\frac{(-q)}{\lambda}\theta'(k)\psi^*(k)\right] \quad \textit{and} \quad d^*(k,q) = \frac{1}{R}\left[\theta(k) + \ln\frac{(-q)^{\lambda}}{\theta'(k)}\right],$$

respectively. Then $J(k,q) = -(-q)^{-\lambda}\exp(-\theta(k))$ is the solution to the HJB equation and the triplet (c^*, d^*, e^*) constitutes the optimal contract. Under the optimal contract (c^*, d^*, e^*), the two state variables evolve according to

$$dk_t = \Big[(k_t+h)e^*(k_t) - c^*(k_t,q_t) - d^*(k_t,q_t) - \delta k_t\Big]dt + \sigma\sqrt{k_t}\,dW_t \tag{3.2}$$

$$\frac{dq_t}{q_t} = \left[\beta - \frac{1}{R}\theta'(k_t)\psi^*(k_t)\right]dt - \frac{\sigma}{\lambda a}\theta'(k_t)e^*(k_t)\psi^*(k_t)\sqrt{k_t}\,dW_t. \tag{3.3}$$

PROOF OF PROPOSITION 2: See Appendix B. □

In general, a two-dimensional optimal control problem like ours gives rise to partial differential equations which are difficult to solve even using a numerical method. We circumvent this difficulty by using the properties of exponential utility. The absence of the wealth effect on the agent's side enables us to derive the second ODE G from the first-order condition $[e]$, meaning that the optimal effort policy is independent of q. On the principal's side, CARA utility not only allows us to conjecture $J(k,q)$ as a multiplicative separable function, but renders the drift of k independent of q; in particular, q does not affect the optimal investment plan. As a result, we can reduce the HJB equation into the uni-dimensional ODE F by canceling out all q terms.

The system of ODEs in Proposition 2 is numerically tractable. Moreover, without the aid of its explicit solutions, we can deduce and establish several properties of the optimal contract directly from the functional forms of F and G. We also prove the existence and uniqueness of a solution to the system in Appendix B.

The following lemma about the asymptotic behavior of $\theta'(k)$ is at the heart of our subsequent results. It asserts that in the optimal contract, the marginal rate of return on investment, $\theta'(k) = -J_k(k,q)/J(k,q)$, converges to zero as firm size rises.

LEMMA 1. *Let $(e^*(k), \theta(k))$ denote the unique solution to the ODEs in Proposition 2, and suppose that $\lim_{k\to\infty} \psi^*(k)$ exists. Then the function $\theta'(k) = G(e^*(k), k)$ converges to zero as $k \to \infty$:*

$$\lim_{k\to\infty} \theta'(k) = 0.$$

PROOF OF LEMMA 1: See Appendix B. □

The following condition is analogous to Assumption 1:

ASSUMPTION 2 (Feller Condition II). $\theta'(0) > R\beta$.

As a counterpart to Assumption 1, Assumption 2 ensures the drift of k_t to be positive for all $k_t > 0$ in the optimal contract, so that the firm does not undergo liquidation almost surely even when its size is very close to zero. The condition is weaker than the sufficient condition for the firm to survive permanently, i.e., $P\{\tau = \infty\} = 1$. Refer to Online Appendix for more details.

Verification

To complete our analysis, we need verify that the conjectured value function is indeed the maximum value the principal can achieve from any incentive-compatible contract. This procedure exploits two lemmas that will appear later (Lemma 2 in the next section and Lemma 3 in Appendix B) and several results obtained from characterizing the system of ODEs in Proposition 2. Thus, readers may find it sufficient to skip the proof of the theorem at the first reading.

THEOREM 1 (Verification). *A solution $J(k, q)$ to the HJB equation provides the principal's value function: Given initial firm size k_0 and the agent's reservation utility q_0, the value the principal can accrue from any incentive-compatible contract is at most $J(k_0, q_0)$.*

PROOF OF THEOREM 1: See Appendix B. □

4. Firm Size and Incentive Provision

In a classic article about dynamic moral hazard, Spear and Srivastava (1987) has shown that the continuation value q_t aggregates all the information on the agent's past performance, and that using q_t as a state variable, the history-dependent optimal contract (Rogerson (1985)) has a simple stationary representation. In a dynamic agency model like ours, which incorporates time-varying firm size, the current firm size k_t plays a role as another state variable that records the firm's past growth path. However, as was pointed out earlier, the corresponding two-dimensional problem often poses an issue of tractability. Most of the recent literature circumvents the issue by taking advantage of the scale invariance principle, and characterizes the optimal contract using the continuation value per capita q_t/k_t as a unique state variable. The variable q_t/k_t can be interpreted as a measure of the agent's stake inside the firm, or a measure of the firm's financial slack in another context (See DeMarzo *et al.* (2012)).

The invariance principle, however, is not applicable to our framework, because the model involves the square-root process k_t and deals with a contract between two *risk-averse* parties. Indeed, the optimal compensation and dividend plans characterized in Proposition 2 rest on both k_t and q_t, which interactively evolve over time in a complicated fashion. This reflects the difficulty of condensing them into a one-dimensional variable. The main feature of our modeling approach, distinct from the previous ones, is that by separating k_t and q_t, we can address the question of how firm size affects the optimal contract. The main goal of the next two sections is to clarify this firm size effect on the corporate decisions about incentive provision and investment.

The Optimal Effort Policy

We begin by studying the effect of firm size on the optimal effort policy e^*. Recall that the agent's optimal choice of effort is invariant to any translation of wealth under the assumption of exponential utility. Hence e^* is unaffected by the continuation value q_t. This property greatly simplifies our analysis of the firm size effect on e^*.[16] It is quite challenging to obtain an explicit form of $e^*(k)$ from the system of ODEs. Even without such an analytical solution, however, we can establish several important properties of $e^*(k)$ from the second ODE $\theta'(k) = G(e^*(k), k)$ and the assumption that $\theta'(k)$ is strictly positive and bounded. All the properties are formally stated in Lemma 2 and displayed in Figure 1.

We first consider the case when firm size is small, in particular, when k_t is close to zero. Note that the size of production shocks $\sigma\sqrt{k_t}\, dW_t$ is assumed to be positively related with k_t. Thus, as the firm shrinks through downsizing, the diffusion of the output process dwindles away. This implies that when k_t is close to zero, $dY_t \approx he_t dt$ provides perfect information about the agent's effort, so that the principal can easily infer e_t from realized output. Accordingly, as firm size approaches zero, the optimal contract implements the first-best effort level ($e^F = a$; see Appendix A).

Aside from this limiting case, the firm's production is exposed to shocks, so the moral hazard problem arises from imperfect monitoring of e_t. This naturally leads to an inefficient level of effort $e^*(k) < a$. Part (a) of Lemma 2 also asserts that for every $k > 0$, the optimal effort policy is bounded below by $e^\dagger(k)$. Here, $e^\dagger(k)$ indicates the minimum level of effort required for the marginal rate of return on investment, $-J_k/J = \theta'(k)$, to remain positive. Hence $e^*(k) > e^\dagger(k)$ implies that the principal encourages the agent to exert a certain level of effort, to the extent that the firm can expand through investment when performance is good. As is noted in footnote 18, the lower bound lies in the interval between 0 and a for every k, although it does not necessarily increase with k.

Part (b) of the lemma describes the optimal effort policy when firm size is sufficiently large. It demonstrates that as firm grows large, the optimal contract retrieves efficiency by implementing the first-best level of effort. Hence $e^*(k)$ is U-shaped as displayed in Figure 1. This result may sound counterintuitive; due to the size-dependent production shock, the provision of incentives gets costly as the firm grows. There is a countervailing firm size effect in our model, however. Recall that the agent's effort determines the expected flow of output $f(k_t, e_t) = (k_t + h)e_t$ which is linear with respect to k_t. Therefore, the signal-to-noise ratio increases with k_t, because the expected output (*signal*) is a function of degree 1 in k whereas the production shock (*noise*) is a function of degree $1/2$.[17] This increased ratio in turn strengthens

[16] If the agent's payoff function is additively separable, the optimal effort varies with q depending on the income effect and the cost of risk premium. In one extreme case in which the agent is very patient, Radner (1985) documented that the efficient level of effort is achieved by the optimal contract and $e^*(k, q)$ decreases with q due to the income effect.

[17] In the previous version, we examined the additive production function $f(k_t, e_t) = k_t + he_t$, which exhibits a constant return to scale. In this case, the signal-to-noise ratio decreases with k, so that $e^*(k)$ approaches a lower bound as k

informativeness of the realized output, thereby reducing the cost of incentive provision.

A similar result - that the agency problem is alleviated as the state variable increases - can be found in He (2009) and the ensuing works, in which firm size follows a geometric Brownian motion and the optimal contract is characterized in terms of the continuation value per capita, q_t/k_t. In comparison with the previous arithmetic Brownian motion model that essentially assumes firm size to be time-invariant, He (2009) demonstrates that the time-varying firm size process provides a *free* incentive by granting the agent a large stake in the firm, equivalently, by increasing q_t/k_t. If the firm's good performance drives q_t/k_t up to a threshold that is the minimum volatility of q_t/k_t for incentive provision, then the agent works voluntarily as he owns enough shares of the firm. The key difference, therefore, lies in the agent's continuation value: *irrespective of* q_t, the optimal contract attains the first-best effort in our model.

(Insert Figure 1 here.)

LEMMA 2. *The optimal effort policy* $e^*(k)$ *satisfies*

(a) $\lim_{k\to 0} e^*(k) = a$ *and* $e^\dagger(k) < e^*(k) < a$ *for every* $k > 0$*, where* $a = e^F$ *is the first-best effort level (See Appendix A) and* $e^\dagger(k)$ *is the largest solution to the following cubic equation:*[18]

$$r(k+h)e^3 - r(k+h)ae^2 + ae = \frac{Ra^2}{R+r}.$$

(b) $\lim_{k\to\infty} e^*(k) = a$*. In addition, if* $\lim_{k\to\infty} \psi^*(k)$ *exists,* $\lim_{k\to\infty} e^{*\prime}(k) = 0$.

PROOF OF LEMMA 2: See Appendix B. □

The results established in Lemma 2 have important implications for the convergence rate of $e^*(k)$ to the efficient level and for the efficiency of the firm's production.

PROPOSITION 3. *The function* $(k+h)(a - e^*(k))$ *is bounded with respect to* k *such that*

$$0 < (k+h)(a - e^*(k)) < \frac{1}{R+r}, \quad \forall\, k \in (0, \infty).$$

Furthermore, suppose that $\lim_{k\to\infty}(k+h)(a - e^*(k))$ *exists and the function* $k^2 e^{*\prime}(k)$ *is bounded. Then* $\lim_{k\to\infty}(k+h)(a - e^*(k)) = \frac{1}{R+r}$ *and* $\lim_{k\to\infty} k e^{*\prime}(k) = 0$.

grows. Hence the square-root process itself is not sufficient to achieve the first-best effort level.

[18] The cubic equation is driven by setting $-\lambda a + (\lambda+1)e\psi(k,e)$, one factor of the denominator of $\theta'(k)$ in (B.5), equal to 0. As the left-hand side of the equation takes on 0 at $e = 0$ but a^2 at $e = a$ and its derivative remains positive for all $e \geq a$ for every $k \geq 0$, it follows by the intermediate value theorem that the largest solution $e^\dagger(k)$ lies between 0 and a.

PROOF OF PROPOSITION 3: The lower and upper bounds are immediate from the inequality $e^{\dagger}(k) < e^{*}(k) < a$ in Lemma 2. For the remaining results, refer to Lemma 4 in Appendix B. □

Note that given k, the expected flow or the drift of output is $(k+h)e^{*}(k)$ in the optimal contract, but $(k+h)a$ in the first-best one. Thus, the function $(k+h)(a-e^{*}(k))$ in Proposition 3, defined by their difference, can be interpreted as a measure of production inefficiency arising from moral hazard. The proposition demonstrates that the function is bounded for all $k > 0$. Its lower and upper bounds are derived from $e^{*}(k) < a$ and $e^{*}(k) > e^{\dagger}(k)$, respectively, and both bounds are independent of firm size.

As the function $(k+h)(a-e^{*}(k))$ is bounded and differentiable for every k, its limit would exist unless $e^{*}(k)$ keeps oscillating between a and $e^{\dagger}(k)$. Proposition 3 says that, except for such a oscillating effort policy, the convergence of $e^{*}(k)$ to the efficient level is fast, of order $1/k$, and thus $\lim_{k\to\infty} ke^{*\prime}(k) = 0$. This implies that as $k \to \infty$,

$$\frac{\partial (k+h)(a-e^{*}(k))}{\partial k} = a - e^{*}(k) + ke^{*\prime}(k) \to 0,$$

because $e^{*}(k) \to a$ by Lemma 2. In other words, the firm improves on production efficiency as it grows, in the sense that there is no distortion in the marginal product of capital.

The Optimal Compensation Scheme

In this subsection we turn to the optimal rate of instantaneous payment in Proposition 2,

$$(4.1) \quad c_t^{*} = \frac{(k_t+h)e^{*}(k_t)^2}{2a} + \frac{1}{r}\ln\left(\frac{\lambda}{\theta'(k_t)\psi^{*}(k_t)}\right) - \frac{1}{r}\ln(-q_t),$$

and discuss the effect of firm size on c_t^{*}. Unlike the effort policy, the payment scheme naturally depends on the continuation value q_t as well; in order to keep her promise with the agent, the principal has to pay for the promised value q_t through the instantaneous payment. To see how c_t^{*} responds to a change in k_t, therefore, we need take into account k_t's direct effect as well as its indirect effect via q_t. From the expression of c_t^{*} above, it is easy to see that a change in k_t directly affects the shape of c_t^{*} through the first two terms. To accommodate the indirect effect, we apply Itô's lemma to the explicit formula of dq_t/q_t in (3.3), and obtain the expression of

$\ln(-q_t)$ as follows:

$$\ln(-q_t) = \ln(-q_0) + \int_0^t \left(\beta - \frac{1}{R}\theta'(k_s)\psi^*(k_s)\right) ds - \int_0^t \frac{\sigma}{\lambda a}\theta'(k_s)e^*(k_s)\psi^*(k_s)\sqrt{k_s}\, dW_s$$
$$- \frac{1}{2}\int_0^t \left(\frac{\sigma}{\lambda a}\theta'(k_s)e^*(k_s)\psi^*(k_s)\sqrt{k_s}\right)^2 ds.$$

Note that q_t keeps track of histories on firm size until time t. Thus, besides the agent's reservation utility $\ln(-q_0)$, it yields the three distinct integral terms, suggesting that the shape of c_t^* is also dependent upon the firm's growth path, $\{k_s\}_{s\in[0,t]}$. Substituting the obtained expression of $\ln(-q_t)$ into (4.1), we can characterize the optimal payment scheme in terms of firm size only.

PROPOSITION 4. *The optimal rate of instantaneous payment can be decomposed into the following six terms:*

$$c^*(k_s,\, s\in[0,t]) = \underbrace{-\frac{1}{r}\ln(-q_0)}_{\text{(i)}} + \underbrace{\frac{(k_t+h)e^*(k_t)^2}{2a}}_{\text{(ii)}} + \underbrace{\int_0^t \frac{\sigma}{Ra}\theta'(k_s)e^*(k_s)\psi^*(k_s)\sqrt{k_s}\, dW_s}_{\text{(iii)}}$$
$$\tag{4.2} + \underbrace{\frac{r}{2}\int_0^t \left(\frac{\sigma}{Ra}\theta'(k_s)e^*(k_s)\psi^*(k_s)\right)^2 k_s\, ds}_{\text{(iv)}}$$
$$+ \underbrace{\frac{1}{r}\ln\left(\frac{\lambda}{\theta'(k_t)\psi^*(k_t)}\right)}_{\text{(v)}} + \underbrace{\frac{1}{r}\int_0^t \left(\frac{1}{R}\theta'(k_s)\psi^*(k_s) - \beta\right) ds.}_{\text{(vi)}}$$

Each term in (4.2) can be interpreted as follows:

(i) reservation utility

(ii) compensation for the effort cost

(iii) compensation risk due to moral hazard

(iv) risk premium due to the compensation risk

(v) adjustment of compensation for future production

(vi) allocation of compensation over time.

The formula (4.2) disentangles the indirect effect of k_t, thereby capturing the exact firm size effect on the optimal payment scheme. In particular, we decompose c_t^* in a similar fashion to the (lump-sum) payment rules in Holmström and Milgrom (1987) and Schättler and Sung (1993), in which the contracting parties have exponential utility like in our model. Hence the formula facilitates comparison with their payment rules (see footnote 19). Apart from the role of firm size, a key difference between their schemes and ours lies in the timing of the payment being made: in those two models, compensation is paid only once at the end of the contract, whereas in our case it is paid continuously over time. As we will explain shortly, these two differences drive the scheme (4.2) to have two distinctive components.

The role of each term from (i) to (iv) is well appreciated. The first two terms provide the agent with his reservation utility and compensation for the cost from exerting $e^*(k_t)$. The stochastic integrand of the term (iii), which is proportional to the volatility of the growth rate of q_t, provides short-term incentives for effort at time t. The term (iv) represents an indemnity to the agent for the risk generated by (iii). Apart from their dependence on firm size, the first four terms also appear in Holmström and Milgrom (1987) and Schättler and Sung (1993).[19]

Before we proceed, it is worth remarking on the incentive term (iii), in particular, concerning how the compensation risk varies with firm size. To this end, we exploit several results established in in Appendix B that as k grows large, $e^*(k) \to a$, $\theta'(k) \to 0$, and $\psi^*(k_t) \to \frac{\lambda}{\lambda+1}$. As a consequence, the integrand of (iii) representing the volatility of dq_t/q_t,

$$\theta'(k_t)e^*(k_t)\psi^*(k_t) \to 0 \quad \text{as } k \to \infty,$$

implying that the agent's continuation value has a stable growth rate in a large firm. This also suggests that the compensation risks for a short-term incentive would be small relative to those in a small firm. The intuition derives from the increasing signal-to-noise ratio over firm size: as k increases, the agent's effort becomes more important relative to the amount of random variation in the level of output. Hence the optimal compensation risk, the weight assigned to the random variation, must decrease with k. As we will prove in the next subsection, the diminishing risk contributes to decreasing pay-performance sensitivity over firm size.

In addition to the four standard components, our framework gives rise to two distinctive terms, (v) and (vi), in the payment scheme. The term (v) establishes a link between the current compensation and the investment plan. Recall that investment corresponds to the firm's free

[19] Writing the optimal lump-sum payment scheme of Holmström and Milgrom (1987) in our notations, we have

$$L_T = -\ln(-q_0) + \int_0^T C(e_t)dt + \int_0^T C'(e_t)dW_t + \frac{r}{2}\int_0^T \sigma^2 C'(e_t)^2 dt,$$

where L_T is the lump-sum payment at the end of the contract, T, and $C(e)$ is the effort cost function. Each term in L_T plays the same role as the term, from (i) to (iv), in (4.2). The way to provide the agent with incentives through the third term of L_T is straightforward: When $C(e)$ is convex, then its derivative $C'(e)$ is an increasing function. Consequently, if the principal wants to implement a higher level of effort, then the wage schedule becomes more volatile. Because firm size is fixed over time and L_T is paid only once at time T in their framework, the terms (v) and (vi) do not appear in L_T.

cash flow. This hints at the possibility that c_t^* is adjusted to supplement investment for growth, which is captured by the term (v). To elaborate on its role, we disregard the indirect effect of k_t on c^* by fixing the level of q_t in (4.1). Note that as $\theta'(k_t) \to 0$ as $k_t \to \infty$, the term (v) is asymptotically positive; that is, there exists a $\bar{k} > 0$ such that $-\ln(\theta'(k_t)\psi^*(k_t)) > 0$ for all $k_t > \bar{k}$. This illustrates the adjustment of payments for investment. The agent in a small firm $(k_t < \bar{k})$ is not fully compensated for the actual effort cost, and the spare amount of money is spent on investment to expand the firm. Instead of deferring the instantaneous payment, the principal promises to pay more later by increasing the drift of q_t. As we will explain shortly, this results in a upward drift of q_t. On the other hand, the agent in a large firm $(k_t > \bar{k})$ is rewarded more than the effort cost, but he is promised a lower continuation value in the future, which results in a downward drift of q_t.

Finally, payment in our model is continuously made over time, so that the principal has to take into account the tradeoff between the immediate compensation c_t^* and the future expected payoffs q_t. The term (vi), whose integrand is the drift part of q_t, captures this tradeoff and represents the allocation of payments over time. In a standard agency model where the agent's utility has income effects, the cost of incentive provision determines the direction of the drift of q_t (Sannikov (2008)). In our model, on the other hand, where the agent's utility has no income effects but the firm can expand or shrink over time, the return on investment determines whether the drift is upward or downward. Just as the return varies with firm size, so does the sign of the drift.[20] To see how it changes over k_t, note that by Lemma 3, when k_t is small, the function $\theta'(k_t)\psi^*(k_t) > R\beta$, implying that q_t has a upward drift. Also, by Lemma 1, when k_t is large, the inequality is reversed so q_t has a downward drift. Therefore, the model predicts that the wage is back-loaded in a small firm but front-loaded in a large firm.

Firm Size and Pay-Performance Sensitivity

It has been a subject prolific of controversy in executive compensation how CEO pay varies relative to changes in firm performance across firm size. The controversy was sparked by a seminal work by Jensen and Murphy (1990), who defined pay-performance sensitivity as the dollar change in CEO wealth per dollar change in firm value and reported that the estimated sensitivity decreases over firm size.[21] In addition to its tiny estimated value ($3.25 increase in

[20] Recall that as the agent has a negative exponential utility function, his continuation value is always negative, too. To avoid any confusions, therefore, it is convenient to multiply both sides by $-q_t$ for the drift of q_t in (3.3). As the resulting drift is $(-q_t)\,[\theta'(k_t)\psi^*(k_t)/R - \beta]$, its sign is determined by the value of $\theta'(k_t)\psi^*(k_t)$.

[21] Schaefer (1998) has argued that the regression model in Jensen and Murphy (1990) $\Delta c_{it} = \gamma_0 + \gamma_c \Delta k_{it}$ implicitly assumes that pay-performance sensitivity γ_c is invariant to firm size. For this reason, he developed a simple econometric

CEO pay per \$1,000 increase in firm value), the decreasing sensitivity with firm size also has been a controversial issue because if the sensitivity is a primary measure of CEO incentives then the empirical result simply implies that the incentives are lower for a large firm.[22]

Although decreasing sensitivity is often attributed to a large standard deviation or weak governance in the market value of a Fortune 500 company, we propose an alternative explanation based on the signal-to-noise ratio scaling with size. Put briefly, with a multiplicative production function, the agent's effort has a bigger impact on the firm's profitability in a large firm, whereas the relative size of the random variation of profitability is small. Hence the agent has an incentive to exert effort even at a low sensitivity. This intuition is very similar to the one in Edmans *et al.* (2009), but there is a major difference in modeling approaches: they employed a (static) talent assignment model of Gabaix and Landier (2008) with moral hazard, but we employ a dynamic contract model.[23]

To derive the sensitivity from our model, we regard the state variable k_t as a measure of the market value of the firm at time t. Then pay-performance sensitivity, denoted as γ_c hereafter, can be approximated by the volatility ratio of c_t^* to k_t:

$$\gamma_c(k_t) \equiv \frac{\Delta c_t^*}{\Delta k_t} = \left.\frac{c_{t+\Delta t}^* - c_t^*}{k_{t+\Delta t} - k_t}\right|_{\Delta t \to 0} = \frac{\text{volatility of } c_t^*}{\text{volatility of } k_t}.$$

As the volatility of k_t is given by $\sigma\sqrt{k_t}$, what is necessary for computation of $\gamma_c(k_t)$ is the volatility of c_t^*, but c_t^* is influenced by both state variables. For a precise measure of relationship between managerial compensation and firm value, therefore, we use the decomposition formula of c_t^* in Proposition 4 to compute the volatility of c_t^*.

To go into details, we apply Itô's lemma only to the terms (ii), (iii), and (v) in (4.2), because the variation arising from the other terms will influence the drift of c_t only. We then divide the resulting expression by the volatility of k_t. This leads us to

$$\gamma_c(k) = \frac{(e^*(k))^2}{2a} + \frac{(k+h)e^*(k)e^{*\prime}(k)}{a} - \frac{1}{r}\left[\frac{\theta''(k)}{\theta'(k)} + \frac{\psi^{*\prime}(k)}{\psi^*(k)} - \frac{1}{\lambda a}\theta'(k)e^*(k)\psi^*(k)\right], \tag{4.3}$$

where $\psi^{*\prime}(k)$ denotes the derivative of the function $\psi^*(k)$ defined in Proposition 2.

model (allowing $\gamma_c(k)$ to vary with k) on the basis of agency theory to show that $\gamma_c(k)$ is proportional to $1/\sqrt{k}$.

[22] $\widehat{\gamma}_c = d\log c_{it}/d\log k_{it}$—referred to as pay-performance *elasticity* in Murphy (1999)—is another prominent measure of the linkage between CEO pay and performance. With regard to its relationship with size, Gibbons and Murphy (1992) reported that, unlike sensitivity, $\widehat{\gamma}_c$ is invariant to firm size. As Murphy (1999) has argued, however, there are pros and cons for each measure. See Baker and Hall (2004) and Edmans *et al.* (2009) for related discussions.

[23] He (2011) also derived a closed-form expression of pay-performance sensitivity in a contract setting with hidden savings, and provided a sufficient and necessary condition for the sensitivity to have a negative relationship with firm size.

The next proposition is then an immediate consequence of Lemma 2 and the results established in Appendix B. Consistent with the empirical prediction, it shows the asymptotically decreasing sensitivity with firm size.

PROPOSITION 5. *Pay-performance sensitivity asymptotically decreases with firm size. That is,*

$$\lim_{k\to\infty} \gamma_c(k) < \lim_{k\to 0} \gamma_c(k).$$

PROOF OF PROPOSITION 5: See Appendix B. □

5. The Optimal Investment Plan

In this section we discuss the effect of firm size on investment. Recall that instantaneous investment $d\mathcal{I}_t$ is determined by the residual cash flow after paying compensations and dividends. Thus, $d\mathcal{I}_t^* = dY_t - (c_t^* + d_t^*)dt$ in the optimal contract. Given an admissible state (k_t, q_t), we take the conditional expectation and define the rate of (expected) investment by

$$I_t^* \equiv \frac{d}{dt}\mathbb{E}[\mathcal{I}_t^* | k_t, q_t] = (k_t + h)e^*(k_t) - c^*(k_t, q_t) - d^*(k_t, q_t),$$

which we will refer to as the optimal investment plan in the sequel.[24] Similarly, we denote by I_t^F the rate of investment in the first-best contract.

The above expression of I_t^* insinuates that, like the remuneration plans c^* and d^*, the investment plan is a function of the two state variables. However, from the explicit formulae of c^* and d^* in Proposition 2, it is readily verified that their sum $c^*(k_t, q_t) + d^*(k_t, q_t)$ is independent of q_t. As the optimal effort policy is dependent upon k_t only, the optimal investment plan is unaffected by q_t. This allows us to write the investment plan as a function of the current firm size, namely, $I_t^* = I^*(k_t)$ and $I_t^F = I^F(k_t)$.

The property of investment being independent of q_t is a key implication of the assumption that the principal's flow utility is exponential. To understand their link, recall that in a classical Merton's portfolio problem, the optimal fraction of wealth being invested in a risky asset is independent of initial wealth when the investor has exponential utility. The investment plan in our model is driven from the same type of portfolio problem. After paying c_t^* to the agent, the

[24] Note that the investment plan I_t^* constitutes the drift of k_t process in the optimal contract, exclusive of the depreciation term.

principal faces a decision problem of allocating the remaining expected cash flow to her own dividends and investment. Here, the continuation value affects the level of expected cash flow or the principal's wealth, but not the rate of return on investment. Therefore, the investment plan I_t^* is determined by k_t, and the same logic also applies to the first-best plan $I_t^F = I^F(k_t)$.

Firm Size and Investment

The above property allows us to address the question of how investment distortions vary with firm size by directly comparing $I^*(k_t)$ and $I^F(k_t)$. It turns out that, in contrast to the prediction of traditional agency models, there could be either under- or over-investment in our model, depending on firm size.

PROPOSITION 6 (Firm Size and Investment Distortions). *When firm size is small, there is under-investment. As the firm grows sufficiently large, however, there is over-investment. More precisely, we have*

$$\lim_{k\to 0} I^F(k) > \lim_{k\to 0} I^*(k) \;\; \textit{and} \;\; \lim_{k\to\infty} I^F(k) < \lim_{k\to\infty} I^*(k).$$

Consequently, the growth rate of a small (large) firm is smaller (larger) for the optimal contract than for the first-best case.

PROOF OF PROPOSITION 6: See Appendix B. □

The underinvestment result relative to the first-best benchmark is straightforward, considering inefficiency due to the moral hazard problem. It is also consistent with what is predicted by other dynamic contracting literature such as He (2009) and DeMarzo *et al.* (2012). On the other hand, because the firm's production improves efficiency by the optimal effort policy converging to the efficient level, it is natural to conjecture that the degree of underinvestment becomes smaller as firm size rises. This is the case in the existing models where investment is increasing in the firm's realized profits. In DeMarzo *et al.* (2012), for instance, the continuation value per capita q_t/k_t, interpreted as a measure of the agent's ownership of company shares or the firm's financial slack, is positively correlated with profits for provision of incentives. In this case, high profits boost investment, thereby increasing firm size as well as q_t/k_t. Consequently, a high continuation value per capita or large firm size mitigates the moral hazard problem, and thus investment approaches the efficient level.

However, our asymptotic result sharply differs from others in that a large firm is prone to over-investment. To approach this result from a different angle, consider the impact of an additional unit of capital on firm value or Tobin's marginal Q. In the optimal contract, marginal Q (denoted MQ^*) is calculated as the partial derivative of firm value with respect to k:

$$MQ^*(k,q) \equiv \frac{\partial \left(J(k,q)+q\right)}{\partial k} = J_k(k,q) = \theta'(k)(-q)^{-\lambda} \exp\left(-\theta(k)\right).$$

Similarly, replacing the principal's value function with $J^F(k,q)$ characterized in Appendix A, we can calculate the first-best marginal Q (denoted MQ^F). To facilitate comparison between MQ^* and MQ^F, we define the ratio of MQ^* to MQ^F by

$$RMQ(k) \equiv \frac{MQ^*(k,q)}{MQ^F(k,q)} = \frac{\theta'(k)}{A_1} \exp(A_1 k + B_1 - \theta(k)).$$

As we have discussed earlier, the agent's continuation value does not affect the firm's investment decision, so we assumed q to be the same in the two regimes when calculating $RMQ(k)$. To examine its asymptotic behavior, we use the result in Lemma 5 that shows $\lim_{k\to\infty} \theta'(k) = 0$. This implies that $\theta(k)$ is increasing at a lower rate than $A_1 k$ for sufficiently large k, so the exponent $A_1 k + B_1 - \theta(k)$ grows large. As a result, $RMQ(k)$ tends to infinity, meaning that an additional unit of capital has a relatively larger impact on firm value in the optimal contract. Therefore, the question of why a large firm is prone to over-investment can be rephrased as follows: Why is MQ^* larger than MQ^F in a large firm?

To address this question, we summarize our previous findings and elaborate on their implications for investment. First, the square-root firm size process results in the optimal effort policy converging to the first-best level, so there is no wedge in the *marginal* return of investment between the two regimes. Put briefly,

$$\frac{\partial (k+h)a}{\partial k} \approx \frac{\partial (k+h)e^*(k)}{\partial k} \quad \text{for a large } k.$$

The lower marginal return on investment in the second-best regime is a primary factor, leading to under-investment in a classic model including ours when firm size is small, in the sense that the principal cannot fully realize the total gain accruing from investment. Second, the contracting parties' utility functions have no wealth effects, so that we can abstract away from the impact of q on investment. The wealth effect on the agent's side typically give rise to under-investment, to the extent that a higher risk premium, resulting from large compensation risks required for incentive provision, decreases the firm's cash flow for undertaking investment.

Excluding the two main sources of investment distortion, what matters from the principal's point of view is the management of risks generated by investment. As the firm's investment increases the variance of production shocks, it is natural to think the risk-averse principal is relatively reluctant to invest when she has to bear larger risks. Put differently, marginal Q would be larger in an environment in which the principal is exposed to smaller risks. However, one bit of conventional wisdom from the contract literature pioneered by Holmström (1979) is that a first-best contract achieves Pareto-optimal risk sharing, whereas a second-best contract trades off risk sharing against provision of incentives. Consequently, in a second-best contract, a portion of risks that the principal has to take for Pareto-optimal risk sharing is passed on to the agent. This leads to larger MQ^* than MQ^F, so that the principal has an incentive to invest more in the optimal or second-best contract.

Risk Sharing in a Large Firm

We attributed over-investment in a large firm to deviation from Pareto-optimal risk sharing. In order to support this argument, we demonstrate that there is indeed such a deviation in the optimal contract. In this subsection, we compute the optimal sensitivity of each contractual term to changes in firm value, and show that it is misaligned with the first-best one. This result partly justifies our argument that the risk-sharing policy in the optimal contract differs from the Pareto-optimal one. The way to compute the sensitivity is very similar to the one we adopted for pay-performance sensitivity in Section 4. So we shall exposit the dividend process here and omit the other details.

Let σ_d^F and σ_d^* denote the volatility of $d^F(k,q)$ and $d^*(k,q)$, respectively. To obtain their explicit formula, we first apply Itô's lemma to the continuation value process of each contract ((A.3) and (3.3)), and then write the dividend process in terms of k only. Applying Itô's lemma again to terms relevant to the volatility, we can explicitly characterize σ_d^F and σ_d^* as a function of k. The sensitivity of d^F and d^*, denoted γ_d^F and γ_d^*, is then calculated as σ_d^F and σ_d^* divided by the volatility of firm value $\sigma\sqrt{k}$, respectively. That is,

$$\gamma_d^F = \frac{A_1}{R(1+\lambda)} \quad \text{and} \quad \gamma_d^*(k) = \frac{1}{R}\left[\theta'(k) - \frac{\theta''(k)}{\theta'(k)} - \frac{1}{a}\theta'(k)e^*(k)\psi^*(k)\right].$$

It follows from Lemma 5 and Corollary 2 that every term in the bracket of $\gamma_d^*(k)$ converges to zero as $k \to \infty$, whereas γ_d^F remains constant over k. Hence $\gamma_d^F > \gamma_d^*(k)$ for sufficiently large k. This result has an implication for dividend smoothing in a dynamic context as well as risk sharing; it shows that compared to d^F, the flow of dividend in a large firm is less sensitive to changes in firm value, so that a relatively low risk is put on the principal.

(Insert Table 1 here.)

Table 1 displays sensitivity of each contractual term and its misalignment with the one from the Pareto-optimal contract. Therefore, Table 1 provides a direct evidence that even in a large firm with a high signal-to-noise ratio, there is still a *moral hazard effect* that deviates from Pareto-optimal risk-sharing for provision of incentives to exert effort. Like the dividend, the agent's compensation c^* turns out to be less sensitive to a change in firm value than the first-best one. On the other hand, the decrease in sensitivity of c^* as well as d^* is backed up by the increase in the sensitivity of investment. The underlying motive for increasing the sensitivity of investment rather than compensation is that such policies will improve the agent's incentives, without the cost of risk premiums and subsequent firm performance.

6. Concluding Remarks

This article studies how firm size affects the optimal contract and investment decision when the size evolves over time with a diminishing volatility. By incorporating a capital accumulation process into a dynamic agency model, the article provides a unified framework where one can explore the impact of the regularity on both moral hazard and investment. The absence of wealth effects due to CARA preference simplifies the optimal contracting problem and enables us to characterize the optimal contract by a system of ordinary differential equations in terms of firm size only. Taking such advantages, we analyzed the impact of firm size on the dynamic incentive and investment.

The diminishing volatility plays two significant roles in our model. First, it results in the increasing signal-to-noise ratio as firm size grows, which in turn leads to improvement on production efficiency in Proposition 3 and a negative relationship between pay-performance sensitivity and size in Proposition 5. Secondly, the regularity results in the decreasing marginal rate of return on investment over size, which is the main driving force behind the downward drift of the continuation value and over-investment in a large firm.

We conclude by making a remark on two important extensions in dynamic contract theory. First, we assumed that the agent has no access to credit markets and is forced to consume what he earns in any period. If the agent can borrow or save freely instead, but if the principal cannot monitor the saving behavior, then hidden saving would distort the optimal intertemporal incentive provision of contracts. As it has been pointed out by Rogerson (1985), the optimal long-term contract must impose a punishment for poor performance so that the agent's marginal utility from savings is always nonnegative. Accordingly, the risk-averse agent is willing to save so as to insure himself against future punishments, and then the principal, anticipating the saving motive, would offer a downside-rigid compensation package. As a result, this distortion in intertemporal incentives would result in the implementable effort policy being below the second-best. The CARA specification is not free from this problem.[25] Furthermore, it is difficult to see how firm size affects the agent's saving motive. Even in a large firm where the moral hazard problem is not severe due to the high signal-to-noise ratio, the downward-rigid compensation scheme may hinder the first-best effort policy from being implemented.

Second, we assumed that the agent is able to commit himself to participation of the contract. Instead, if the agent could quit in any period, we need additional participation constraints at each possible history. To see how the constraints affect our results, note that in the optimal contracts, good performance must result in an increase in both q_t and k_t. Hence the two state variables must be positively correlated, implying that the limited commitment issue would not be problematic in a large firm where the participation constraints are likely to be slack. In a small firm, on the other hand, the constraints are more likely to bind. Then from the principal's vantage point, it is difficult to provide proper incentives through q_t, as q_t is now downward-protected for all t. So the optimal effort policy proposed in the article would not be implementable. Furthermore, it is also expected that the proper incentive has to be provided through c_t rather than q_t, thereby leading to a higher pay-performance sensitivity in a small firm compared to the full commitment case.

Appendices

A. First-Best Contract

In this section we present an explicit solution to the contract problem with full information. It turns out that the first-best contract can be derived in a very similar way to the second-best contract. We can thus bypass a detailed discussion of the verification procedure.

[25] In the early dynamic moral hazard models without firm dynamics (for instance, Fudenberg *et al.* (1990)), the CARA preference helps to simplify the agent's dynamic saving problem into a static problem, to the extent that the agent's saving motive is independent of the past history—the agent's continuation value or wealth. However, this is not the case in our model, because the model involves another state variable (firm size) and it is difficult to pin down its relationship with the saving decision.

When the agent's action is perfectly observable, the volatility term Γ of the agent's continuation value process is to be chosen by the principal as she is free from the incentive provision issue. We can therefore reformulate the principal's problem into the following HJB equation:

$$\begin{aligned}\beta J(k,q) = \max_{c,d,e,\Gamma} \Big\{ & v(d) + J_k[f(k,e) - c - d - \delta k] + J_q\left[\beta q - u(c,e)\right] \\ & + \frac{\sigma^2 k}{2}\left(J_{kk} + 2J_{kq}\Gamma + J_{qq}\Gamma^2\right)\Big\}.\end{aligned} \tag{A.1}$$

The only difference from the HJB equation for second-best optimality is that Γ is now another choice variable for the principal and is used for maximizing her own value. Put differently, Γ is determined so as to achieve Pareto-optimal risk sharing between the two risk-averse parties.

Assuming the CARA utility and the production technology described in Section 3, we can explicitly solve the HJB equation (A.1) as follows:

PROPOSITION 7. *The first-best contract, denoted (c^F, d^F, e^F), is characterized by $e^F = a$,*

$$c^F(k,q) = \frac{(k+h)a}{2} - \frac{1}{r}\ln\left(-\frac{qA_1}{\lambda}\right), \quad \textit{and}$$
$$d^F(k,q) = \frac{1}{R}\left[A_1 k + B_1 + \lambda \ln(-q) - \ln A_1\right],$$

where A_1 and B_1 are constants:

$$A_1 = \frac{\frac{a}{2} - \delta}{\frac{\sigma^2}{2(\lambda+1)} + \frac{1}{R}} \quad \textit{and} \quad B_1 = \frac{Rha}{2} - \lambda\ln\lambda + (1+\lambda)\left[\frac{R\beta}{A_1} - 1 + \ln A_1\right].$$

PROOF OF PROPOSITION 7: The first-order condition with respect to each choice variable is

$$[\Gamma]: \quad \Gamma = -\frac{J_{kq}}{J_{qq}};$$
$$[c]: \quad -J_k - J_q \exp\left(-r\left(c - \frac{(k+h)e^2}{2a}\right)\right) = 0;$$
$$[d]: \quad \exp\left(-Rd\right) - J_k = 0; \text{ and}$$
$$[e]: \quad J_k(k+h) + J_q\frac{(k+h)e}{a}\exp\left(-r\left(c - \frac{(k+h)e^2}{2a}\right)\right) = 0.$$

Like the second-best one, we conjecture that the principal's value function is of the form $J(k,q) = -(-q)^{-\lambda}\exp(-\theta(k))$, where $\theta : [0,\infty) \to \mathbb{R}$ is a C^2 function and $\lambda = R/r$. Just as in Section 3, we maintain the assumption of θ that $\theta'(k) \in (0,\infty)$ and $(\theta'(k))^2 > (\lambda+1)\theta''(k)$ to ensure that the conjectured value function is concave in the two state variables.

First, the (constant) first-best effort $e^F = a$ is immediate from taking ratios of the conditions [c] to [e] regardless of any functional forms of J. The first 3 conditions along with our guess determine the optimal volatility term of q, the rate of consumption, and the rate of dividend in order:

$$\Gamma^F = \frac{(-q)\theta'(k)}{\lambda+1}, \quad c^F = \frac{(k+h)a}{2} - \frac{1}{r}\ln\left(\frac{(-q)\theta'(k)}{\lambda}\right), \quad d^F = \frac{1}{R}\left[\theta(k) + \ln\left(\frac{(-q)^\lambda}{\theta'(k)}\right)\right].$$

We use (c^F, d^F, e^F) to characterize the drift of the capital process, exclusive of its depreciation term, into a function of k only:

$$I^F(k) \equiv ke^F + he^F - c^F - d^F = \frac{(k+h)a}{2} - \frac{\theta(k)}{R} + \frac{\lambda+1}{R}\ln\theta'(k) - \frac{\ln\lambda}{r}.$$

Substituting the optimal policies back into the equation (A.1) delivers the following nonlinear ordinary differential equation (ODE) of $\theta(k)$:

$$(1+\lambda)\beta = \theta'(k)\left[\frac{\lambda+1}{R} - I^F(k) + \delta k\right] + \frac{\sigma^2 k}{2(\lambda+1)}\left(\theta'(k)\right)^2 - \frac{\sigma^2 k}{2}\theta''(k).$$

Due to the $\ln(\theta'(k))$ term in $I^F(k)$, it is natural to put $\theta(k)$ as a linear function of k. Plugging $\theta(k) = A_1 k + B_1$ into the ODE above and then solving for the two constants A_1 and B_1, we can explicitly characterize them as follows:

$$A_1 = \frac{\frac{a}{2} - \delta}{\frac{\sigma^2}{2(\lambda+1)} + \frac{1}{R}} \quad \text{and} \quad B_1 = \frac{Rha}{2} - \lambda\ln\lambda + (1+\lambda)\left[\frac{R\beta}{A_1} - 1 + \ln A_1\right]. \quad \square$$

Proposition 7 also shows us how the two state variables evolve over time.

COROLLARY 1. *The capital $\{k_t\}_{t\in[0,\infty)}$ and the agent's continuation value $\{q_t\}_{t\in[0,\infty)}$ processes in the first-best contract evolve as follows:*

$$\text{(A.2)} \quad dk_t = \left[\left(\frac{a}{2} - \delta - \frac{A_1}{R}\right)k_t - (1+\lambda)\left(\frac{\beta}{A_1} - \frac{1}{R}\right)\right]dt + \sigma\sqrt{k_t}\,dW_t$$

$$\text{(A.3)} \quad \frac{dq_t}{q_t} = \left(\beta - \frac{A_1}{R}\right)dt - \frac{A_1}{\lambda+1}\sigma\sqrt{k_t}\,dW_t.$$

To keep k_t positive and its process well-defined over time, we make assumptions about the primitives so that the drift term of k_t remains positive for all k_t.[26] This is a version of the Feller

[26]The coefficient of k_t in the drift term $\frac{a}{2} - \delta - \frac{A_1}{R}$ is always positive for all feasible primitives.

condition we remarked on in Assumption 1:

$$\frac{\beta}{A_1} - \frac{1}{R} \leq 0 \Rightarrow A_1 \geq \beta R.$$

B. Omitted Proofs

Proofs of Proposition 1

Given a contract $(c, d, e) \in \mathcal{S}$, consider the agent's expected payoff evaluated at time t,

$$\hat{U}_t = \int_0^t e^{-\beta s} u(c_s, \hat{e}_s) ds + e^{-\beta t} q_t(c, d, e),$$

when the agent chooses an alternative level of effort $\hat{e}$ up to t and follows the recommended level of effort e from t on. Differentiating $\hat{U}_t$ with respect to t gives

$$\text{(B.1)} \quad d\hat{U}_t = e^{-\beta t} \left[u(c_t, \hat{e}_t) + \Gamma_t f(k_t, \hat{e}_t) - u(c_t, e_t) - \Gamma_t f(k_t, e_t) \right] dt + e^{-\beta t} \Gamma_t \sigma \sqrt{k_t}\, dW_t^{\hat{e}},$$

where we used the dynamics of q_t in (2.3) for simplifying $d(e^{-\beta t} q_t)$ and the relationship between W_t^e under P^e and $W_t^{\hat{e}}$ under $P^{\hat{e}}$,

$$\sigma \sqrt{k_t}\, dW_t^e = \sigma \sqrt{k_t}\, dW_t^{\hat{e}} + (f(k_t, \hat{e}_t) - f(k_t, e_t)) dt.$$

For necessity, suppose to the contrary that (2.4) does not hold on a set of positive measure. Then it follows from (B.1) that $\hat{U}_t$ has a positive drift by choosing $\hat{e}_t$ to maximize $u(c_t, \hat{e}_t) + \Gamma_t f(k_t, \hat{e}_t)$ and thus $\mathbb{E}^{\hat{e}}[\hat{U}_t] > \hat{U}_0 = q_0(c, d, e)$, violating the (IC) condition. For sufficiency, suppose (2.4) holds. Then the drift of $\hat{U}_t$ becomes negative, meaning that the process is a supermartingale for every deviation $\hat{e}$. Therefore, we have $q_0(c, d, e) \geq \mathbb{E}^{\hat{e}}[\hat{U}_\infty] = q_0(c, d, \hat{e})$.[27]

□

Proofs of Proposition 2

We begin with the volatility of the agent's continuation value, Γ_t. The multiplicative production technology and the CARA preferences allow us to pinpoint Γ_t necessary for implementation

[27] The proof is basically the same as the proof of Proposition 2 in Sannikov (2008), but we provide our own proof for a self-contained article. A more general proof can be found in Proposition 5.1 in Williams (2013) or Theorem 4.2 in Schättler and Sung (1993).

of an instruction e_t. To see this, recall from Proposition 1 that, for the instruction to be self-enforced, e_t should maximize the objective function $u(c_t, e) + \Gamma_t f(k_t, e)$, which is now globally concave in e. Thus, Γ_t is uniquely determined by the following first-order condition:

$$\Gamma_t = -\frac{u_e(c_t, e_t)}{f_e(k_t, e_t)} = -\frac{re_t}{a} u(c_t, e_t),$$

which suggests that the volatility can be written as a function of c_t and e_t; so let us write $\Gamma_t = \Gamma(c_t, e_t)$. Suppressing the time subscript, the relationship between Γ and u driven by the condition can also be used to simplify the first-order partial derivatives of Γ and u as follows:

$$\text{(B.2)} \quad \Gamma_c(c, e) = -r\Gamma, \quad \Gamma_e(c, e) = \frac{a + r(k+h)e^2}{ae}\Gamma, \quad u_c(c, e) = \frac{a\Gamma}{e}, \quad u_e(c, e) = -(k+h)\Gamma.$$

We then take the derivative of the HJB equation with respect to each contractual term, identify the first-order conditions, and use (B.2) to simplify the conditions into

$$[c]: \quad -J_k - J_q \frac{a\Gamma}{e} - \left(J_{kq}\Gamma + J_{qq}\Gamma^2\right) r\sigma^2 k = 0$$

$$[d]: \quad \exp(-Rd) - J_k = 0$$

$$[e]: \quad J_k(k+h) + J_q(k+h)\Gamma + \left(J_{kq}\Gamma + J_{qq}\Gamma^2\right) \frac{a + r(k+h)e^2}{ae}\sigma^2 k = 0.$$

From the first-order conditions for $[c]$ and $[e]$, we obtain the following alternative expression of Γ:

$$\text{(B.3)} \quad \Gamma = -\frac{J_k}{J_q}\left[1 + \frac{r(k+h)}{a}e(e-a)\right]\frac{e}{a} = \frac{e}{\lambda a}(-q)\theta'(k)\psi(k, e),$$

where $\psi(k, e) \equiv 1 + \frac{r(k+h)}{a}e(e-a)$ represents the expression in the bracket above. Equating the two expressions of Γ and then solving for c, we obtain the flow of consumption

$$c(k, q) = \frac{(k+h)e^2}{2a} - \frac{1}{r}\ln\left(\frac{1}{\lambda}(-q)\theta'(k)\psi(k, e)\right).$$

Solving the remaining first-order condition $[d]$ gives the flow of dividend,

$$d(k, q) = \frac{1}{R}\left[\theta(k) + \lambda\ln(-q) - \ln\theta'(k)\right].$$

We then use (c, d) to write the drift term of the capital evolution process, exclusive of capital depreciation, as

$$\text{(B.4)}\quad I(k,e) = (k+h)e - \frac{(k+h)e^2}{2a} + \frac{1}{r}\ln\left(\frac{\theta'(k)\psi(k,e)}{\lambda}\right) - \frac{1}{R}\left[\theta(k) - \ln\theta'(k)\right].$$

Notice that the function $I(k,e)$ which we will refer to as the investment policy in Section 5, is independent of the agent's continuation payoff.

Now we derive the system of ODEs for the remaining two functions, e and $\theta(k)$, which completely characterizes the optimal contract together with (c,d,I,Γ) specified above. For $\theta(k)$ in the conjectured value function, we substitute Γ in (B.3) into the first-order condition [e] and solve for $\theta'(k)$ to obtain

$$\text{(B.5)}\quad \theta'(k) = \frac{a(k+h)\lambda(a-e)}{\sigma^2 k\psi(k,e)[-\lambda a + (\lambda+1)e\psi(k,e)]} \equiv G(e,k)$$

Note that the expression on the right side is a function of e and k so we label it by $G(e,k)$. Also, the equation (B.5) suggests that the optimal effort is a function of k only.

Denoting by $e(k)$ the optimal effort policy, we substitute the above (c,d,Γ) into the HJB equation to obtain the following second-order ODE:

$$\text{(B.6)}\quad \begin{aligned}(1+\lambda)\beta = {} & \theta'(k)\left[\frac{1}{R} - I(k) + \delta k + \frac{\psi(k,e(k))}{r}\right] \\ & + \frac{\sigma^2 k}{2}\left(\theta'(k)\right)^2\left[1 - 2\frac{e(k)\psi(k,e(k))}{a} + \frac{\lambda+1}{\lambda}\left(\frac{e(k)\psi(k,e(k))}{a}\right)^2\right] - \frac{\sigma^2 k}{2}\theta''(k),\end{aligned}$$

where the investment policy $I(k) = I(e(k),k)$ reduces to a function of k after substitution $e = e(k)$. Rearranging and substituting $\theta'(k) = G(e(k),k)$ into (B.6) yields

$$\text{(B.7)}\quad \begin{aligned}k\theta''(k) = {} & \frac{2}{\sigma^2}G(e(k),k)\left[\frac{1}{R} - I(k) + \delta k + \frac{\psi(k,e)}{r}\right] \\ & + kG^2(e(k),k)\left[1 - 2\frac{e(k)\psi(k,e(k))}{a} + \frac{\lambda+1}{\lambda}\left(\frac{e(k)\psi(k,e(k))}{a}\right)^2\right] - \frac{2(1+\lambda)\beta}{\sigma^2} \\ \equiv {} & H(\theta(k),e(k),k).\end{aligned}$$

Note that the right-hand side of (B.7) is a function of $\theta(k)$, $e(k)$ and k, and thus we will label it by $H(\theta(k),e(k),k)$. Now we take the derivative of both sides of (B.5) with respect to k to obtain

$$\text{(B.8)}\quad \theta''(k) = G_e(e(k),k)e'(k) + G_k(e(k),k).$$

Then the desired second ODE follows from (B.7) and (B.8):

$$e'(k) = \frac{1}{G_e(e(k),k)} \left[\frac{H(\theta(k), e(k), k)}{k} - G_k(e(k),k) \right] \equiv F(\theta(k), e(k), k).$$

In summary, the system of ODEs is given by

$$\begin{cases} \theta'(k) = G(e(k),k) \\ e'(k) = F(\theta(k), e(k), k) \end{cases}$$

with the boundary condition $(e(0), \theta(0)) = \left(a, \ln\left[R\beta(r\beta)^{\lambda}\right]\right)$. The first boundary condition follows from Lemma 2, and the second from solving (3.1) for $\theta(0)$. We discuss the existence and uniqueness of their solutions in Appendix B.

Denote by $e^*(k)$ the solution to the system of ODEs. Substituting $e = e^*(k)$ into the functional forms of c and d above leads to the optimal consumption and dividend policies, $c^*(k,q)$ and $d^*(k,q)$. Lastly, substituting (c^*, d^*, e^*) into the k- and q-process provides the evolution of the two state variables under the optimal contract. The existence and uniqueness of (k,q) processes satisfying the system of stochastic differential equations (3.2) and (3.3) are established by Yamada and Watanabe (1971), as $\sqrt{k}$ is Hölder-continuous with exponent $1/2$. The proof is now complete. □

Existence and Uniqueness of the System of ODEs in Proposition 2

Define by X a set of $(e(k), \theta(k))$ such that $e(k) \leq a$ for all k and $\theta(k)$ has a bounded derivative. As we prove in Lemma 2 below, both $e^*(k)$ and $\psi^*(k)$ are uniformly bounded, which guarantees that $G(e^*(k), k)$ and $\theta'(k)$ are also uniformly bounded. Consequently, we can restrict ourselves to the set X for finding a solution to the ODEs. In addition, note that the two functions F and G consisting of ODEs in are continuously differentiable and do not explode.

Choose $k_0 > 0$ as an initial value of capital and let $(\theta(k_0), e(k_0))$ be the corresponding initial value condition. By the standard theory on the system of first order non-linear ordinary differential equations, there exists a unique C^1 solution $(\theta(k), e^*(k))$ in some neighborhood of k_0. This local existence and uniqueness result is readily extended to any arbitrary finite capital level $k_M < \infty$ so long as the solution is bounded as $k \to k_M$, which is the case in our problem. In other words, the solution uniquely exists for $k \in [k_0, k_M]$.

Finally, the above argument is easily extended to the case where $k_0 \to 0$, because in our problem the boundary values are in fact derived by calculating the limit of the solutions. More specifically, for the effort policy, $\lim_{k_0 \downarrow 0} e^*(k_0) = a$ as is shown in Lemma 2. For the initial value of $\theta(k)$, given any $k_0 > 0$, we can compute $J(k_0, q)$ and thus $\theta(k_0)$ by assuming that the firm is liquidated at k_0, like we defined $\theta(0)$ in Section 3. Then the desired initial value is simply driven by taking the limit: $\lim_{k_0 \downarrow 0} \theta(k_0) = \theta(0)$. $\square$

Proof of Theorem 1

Suppose that $J(k, q)$ is a solution to the HJB equation. For every incentive-compatible contract (c, d, e), we define the principal's auxiliary gain process $V = \{V_t\}_{t \in [0,\infty)}$ as

$$V_t(c, d, e) \equiv \int_0^t e^{-\beta s} v(d_s) ds + e^{-\beta t} J(k_t, q_t).$$

Here k_t and q_t are the two state variables at time t induced by (c, d, e), and hence V_t represents the principal's expected total payoffs (evaluated at time t) when she offered the contract (c, d, e) until time t but plans to offer the optimal contract (c^*, d^*, e^*) afterwards. We now show that the process V is a super-martingale, but is a martingale when the contract (c, d, e) is optimal.

Using Itô's lemma, we compute the differential of V:

$$dV_t = e^{-\beta t} A_t dt + \beta e^{-\beta t} \sigma \sqrt{k_t} \left\{ J_k - q_t J_q \frac{1}{\lambda a} \theta'(k_t) e(k_t) \psi(k_t, e(k_t)) \right\} dW_t$$

where the drift term A_t is

$$A_t \equiv v(d_t) - \beta J + J_k \left[I(k_t, q_t) - \delta k_t \right] + J_q \left[\beta q_t - u(c_t, e_t) \right] + \frac{\sigma^2 k}{2} \left[J_{kk} + 2 J_{kq} \Gamma_t + J_{qq} \Gamma_t^2 \right].$$

Then it follows by definition of the HJB equation that $A_t \leq 0$ for every incentive-compatible contract and $A_t = 0$ for the optimal contract. It remains to show that the diffusion of V is square-integrable *in the optimal contract*. To this end, we substitute the computed derivatives (J_k and J_q) into the diffusion term of V_t and rewrite the term as

$$\beta \sigma \exp(-\theta(k_t)) \theta'(k_t) \left(1 - \frac{1}{a} e(k_t) \psi(k_t, e(k_t)) \right) e^{-\beta t} \frac{\sqrt{k_t}}{(-q_t)^{\lambda}}.$$

We prove in Lemma 2 and 5 that $\theta'(k_t)$, $e(k_t)$, and $\psi(k_t, e(k_t))$ are all bounded. Hence it suffices to show that the remaining term $\frac{e^{-\beta t} \sqrt{k_t}}{(-q_t)^{\lambda}}$ is square-integrable. This is immediate, however, from the fact that the agent's continuation value q_t is bounded above by 0 for any time t; $q_t = 0$ under the CARA preferences implies that the rate of payment must be infinite after time t with positive probability, but this is never feasible. □

Proof of Lemma 2

Suppose that a continuously differentiable effort policy $e^*(k)$ solves the ODE $\theta'(k) = G(e^*(k), k)$ in Proposition 2, which takes a form of

$$\text{(B.9)} \quad \theta'(k) = \frac{a(k+h)\lambda(a - e^*(k))}{\sigma^2 k\psi(k, e^*(k))\left[-\lambda a + (\lambda+1)e^*(k)\psi(k, e^*(k))\right]}.$$

To prove $\lim_{k\to 0^+} e^*(k) = a$, note that the denominator of (B.9) converges to zero as $k \to 0^+$ due to the factor k. For $\theta'(k)$ to be bounded, therefore, the factor $a - e^*(k)$ on the numerator must be zero, as $\theta'(k)$ would otherwise diverge.

To prove that the optimal effort is below the first-best level, first observe that the function $\psi(k,e) = 1$ at $e = a$ irrespective of k. This implies that the denominator is positive in a neighborhood of $k = 0$; more precisely, there exists an $\epsilon > 0$ such that for every $k \in (0, \epsilon)$, the denominator is strictly positive, because by the continuity of $e^*(k)$,

$$\psi(k, e^*(k))\left[-\lambda a + (\lambda+1)e^*(k)\psi(k, e^*(k))\right] \approx a > 0.$$

To obtain $\theta'(k) > 0$ in this neighborhood, therefore, the numerator of (B.9) must be strictly positive, leading to $e^*(k) < a$ for $k \in (0, \epsilon)$. Furthermore, $e^*(k) < a$ is readily extended to all $k > 0$, for otherwise $\theta'(k) < 0$ for some $k > 0$ by the intermediate value theorem.[28] Hence we establish $e^*(k) < a$ for all $k > 0$.

To prove that $e^*(k)$ is bounded below by $e^\dagger(k)$, we define a function $g(k,e) \equiv -\lambda a + (\lambda + 1)e\psi(k,e)$ on the space (k,e). Then $e^\dagger(k)$, defined in the statement of Lemma 2, is simply the largest solution to the equation $g(k,\cdot) = 0$. With this in mind, we first shall prove $e^\dagger(k) < a$ by contradiction. Suppose $e^\dagger(k) \geq a$. Then it follows from $\psi_e(k,e) > 0$ for all $e \geq a$ and $g(k, e^\dagger(k)) = 0$ that

$$\frac{\lambda a}{(\lambda+1)e^\dagger(k)} = \psi(k, e^\dagger(k)) \geq \psi(k,a) = 1 \quad \text{or} \quad a \geq \frac{(\lambda+1)e^\dagger(k)}{\lambda},$$

leading to a contradiction. For the remaining part, note that as $e^\dagger(k)$ is the ”largest” solution, $g(k,e) > 0$ for all $e > e^\dagger(k)$ and thus $\psi(k,e) > 0$. This implies that the whole denominator of

[28] Suppose to the contrary that $e^*(k) > a$ for some k. Then it follows from continuity of $e^*(k)$ and $\psi(k, e^*(k))$ that there exists a $k^o \in (0,k)$ such that (1) $e^*(k^o) > a$ but very close to a and (2) $\psi(k^o, e^*(k^o))$ is close to one. Because the derivative $\theta'(k)$ will have a negative numerator but a positive denominator for such k^o, $e^*(k) > a$ results in a contradiction.

(B.9) changes its sign from negative to positive at $e = e^{\dagger}(k)$. Hence for a positive $\theta'(k)$, $e^*(k)$ must lie in the interval between $e^{\dagger}(k)$ and a. This completes the proof of part (a).

(Insert Figure 2 here.)

For part (b), note that $e^{\dagger}(k) < e^*(k) < a$ implies

$$\frac{\lambda}{\lambda+1} < \psi(k, e^{\dagger}(k)) < \psi(k, e^*(k)) < \psi(k, a) = 1 \quad \forall\, k > 0,$$

that is, the function $\psi^*(k) \equiv \psi(k, e^*(k))$ is bounded. Refer to Figure 2. Taking the limit of $\psi^*(k)$ as $k \to \infty$ gives

$$\frac{\lambda}{\lambda+1} \leq \lim_{k\to\infty} \psi^*(k) = \lim_{k\to\infty} \left[1 + \frac{r(k+h)}{a} e^*(k)\,(e^*(k) - a)\right] \leq 1.$$

For $\psi^*(k)$ to be bounded, however, it must be the case that $\lim_{k\to\infty} e^*(k) = a$, as otherwise the limit would be unbounded below.[29]

Finally, to prove $\lim_{k\to\infty} e^{*\prime}(k) = 0$, consider the ODE $e^{*\prime}(k) = F(\theta(k), e^*(k), k) = \frac{\theta''(k) - G_k(e^*(k),k)}{G_e(e^*(k),k)}$. Notice that if $\lim_{k\to\infty} \psi^*(k)$ exists, then its value must be $\frac{\lambda}{\lambda+1}$ as is shown in Lemma 4. Exploiting this result, a little bit of algebra shows that $\lim_{k\to\infty} G_e(e(k), k) = -\infty$ whereas the numerator in the ODE is bounded. This completes the proof. $\square$

Technical Lemmas

In this subsection, we establish several lemmas that articulate the limiting behavior of $\theta'(k)$ and other functions when k is sufficiently small and large, respectively. The lemmas are frequently used for the proof of the ensuing results.

LEMMA 3. $\theta'(0) \equiv \lim_{k\to 0^+} \theta'(k) \leq A_1$.

PROOF OF LEMMA 3: Define a function $\phi : (0, \infty) \to \Re$ as

$$\phi(x) = \frac{Rha}{2} - \lambda \ln \lambda + (1+\lambda)\left[\frac{\beta R}{x} + \ln x - 1\right].$$

The function $\phi(x)$ is derived from the equation (B.7) by substituting $k = 0$ in (B.7) and then solving for $\theta(0)$. As a result, $\phi(\theta'(0)) = \theta(0)$ follows by definition and $\phi(A_1) = B_1$ by straightforward computation, where the two constants A_1 and B_1 are from the first-best contract char-

[29] Another way to prove $\lim_{k\to\infty} e^*(k) = a$ is to use the sandwich theorem; because $e^{\dagger}(k) < e^*(k) < a$ and $e^{\dagger}(k) \to a$ in the limit as $k \to \infty$, $e^*(k)$ must converge to a as well. However, we show in Lemma 4 that as k grows, $e^*(k)$ is getting closer to $e^{\dagger}(k)$ rather than to a, implying that $\lim_{k\to\infty} \psi(e^*(k), k) = \lim_{k\to\infty} \psi(e^{\dagger}(k), k) = \frac{\lambda}{\lambda+1} < 1 = \lim_{k\to\infty} \psi(a, k)$.

acterized in Proposition 7. Furthermore, it can be readily checked that (i) $\phi(x)$ is decreasing on $(0, R\beta)$ but increasing on $(R\beta, \infty)$, so $\phi(x)$ takes on a global minimum value at $x = R\beta$; and that (ii) $\lim_{x\to 0^+} \phi(x) = \lim_{x\to\infty} \phi(x) = \infty$. Refer to Figure 3.

Note that $\theta(0) \le B_1$, as $J^F(0,q) \ge J^*(0,q)$ for every admissible q. But as A_1 and $\theta'(0)$ are larger than $R\beta$ by Assumption 1 and 2, both of them lie in the region where $\phi(x)$ is increasing. Therefore, $\theta(0) \le B_1$ leads to $\theta'(0) \le A_1$. □

(Insert Figure 3.)

LEMMA 4. *Suppose that* $\lim_{k\to\infty} \psi^*(k)$ *exists. Then* $\lim_{k\to\infty} \psi^*(k) = \frac{\lambda}{\lambda+1}$.[30]

PROOF OF LEMMA 4: Recall that in the proof of Lemma 2, we proved $\frac{\lambda}{\lambda+1} < \psi^*(k) < 1$ for all k. This inequality allows us to subdivide the set of possible values of $\lim_{k\to\infty} \psi^*(k)$ into three cases: (i) $\lim_{k\to\infty} \psi^*(k) \in \left(\frac{\lambda}{\lambda+1}, 1\right)$; (ii) $\lim_{k\to\infty} \psi^*(k) = 1$; and (iii) $\lim_{k\to\infty} \psi^*(k) = \frac{\lambda}{\lambda+1}$. Below it is shown by contradiction that (i) and (ii) are not true. Before proceeding, notice that by (B.9) we have $\lim_{k\to\infty} \theta'(k) = 0$ in both (i) and (ii). Also, it is immediate from (B.5) that $\lim_{k\to\infty} k\theta'(k) > 0$ in case (i) whereas $\lim_{k\to\infty} k\theta'(k) = 0$ in case (ii).

CASE 1: Suppose that $\lim_{k\to\infty} \psi^*(k) \in \left(\frac{\lambda}{\lambda+1}, 1\right)$. On account of heavy algebra works, we shall omit the details and sketch out the way to derive a contradiction.[31] We first demonstrate $\theta''(k) \to 0$. For this purpose, we rewrite the ODE $e^{*\prime}(k) = F(\theta(k), e^*(k), k)$ as

$$\text{(B.10)} \quad e^{*\prime}(k) = \frac{1}{G_e(e^*(k),k)}\Big[\theta''(k) - G_k(e^*(k),k)\Big],$$

where we used $H(\theta(k), e^*(k), k) = k\theta''(k)$ from (B.7) to create a link between $\theta''(k)$ and $e^{*\prime}(k)$. From the explicit forms of G_k and G_e which can be computed from (B.5), it can be shown that as $k \to \infty$, both G_k and kG_k converge to zero but G_e converges to a negative value. Then, in order to satisfy (B.10), $\theta''(k)$ must converge to zero, because the left-hand side of (B.10) tends to zero by Lemma 2.

Next, we substitute the optimal investment plan $I^*(k) = I(k, e^*(k))$ (Refer to (B.4)) into the ODE (B.6) and then take the limit of both sides as $k \to \infty$. This gives us

$$\lim_{k\to\infty} k\theta''(k) = -\frac{2}{\sigma^2}\left[(1+\lambda)\beta + (a/2 - \delta)\lim_{k\to\infty} k\theta'(k)\right] < 0,$$

[30]This result can be established with or without the assumption of $k^2 e^{*\prime}(k)$ being bounded. When we assume $k^2 e^{*\prime}(k)$ to be bounded, the proof becomes a bit more succinct. But we do not rely on this assumption to emphasize that the desired result holds in either case.

[31]The exact proof is available from the authors upon request.

because $\lim_{k\to\infty} k\theta'(k)$ in the bracket takes a positive value and $a > 2\delta$ by Assumption 1. On the other hand, multiplying both sides of (B.10) by k and taking the limit, we have

$$\text{(B.11)}\quad \lim_{k\to\infty} ke^{*\prime}(k) = \left(\lim_{k\to\infty} \frac{1}{G_e(e^*(k), k)}\right)\left(\lim_{k\to\infty} k\theta''(k)\right) > 0$$

because both limits on the right-hand side are negative. (B.11) results in $\lim_{k\to\infty} \psi^{*\prime}(k) > 0$, where $\psi^{*\prime}$ represents the total derivative of $\psi^*(k)$ with respect to k:

$$\text{(B.12)}\quad \psi^{*\prime} = \frac{d\psi(k, e^*(k))}{dk} = \psi_k + \psi_e e^{*\prime}(k) = \frac{r}{a}\left[e^*(k)(e^*(k) - a) + (2e^*(k) - a)(k+h)e^{*\prime}(k)\right].$$

Finally, we apply the natural logarithm to (B.9) and then take the derivative of both sides with respect to k, to obtain

$$\text{(B.13)}\quad \frac{\theta''(k)}{\theta'(k)} = \underbrace{\frac{1}{k+h} - \frac{1}{k}}_{\text{(i)}} - \underbrace{\frac{e^{*\prime}(k)}{a - e^*(k)}}_{\text{(ii)}} - \underbrace{\frac{\psi^{*\prime}}{\psi^*}}_{\text{(iii)}} - \underbrace{(\lambda+1)\frac{\psi^{*\prime}e^*(k) + \psi^* e^{*\prime}(k)}{-\lambda a + (\lambda+1)\psi^* e^*(k)}}_{\text{(iv)}}$$

Note that although the term (i) clearly approaches zero, the other terms approach positive constants. As a result, we have $\lim_{k\to\infty} \theta''(k)/\theta'(k) < 0$. However, this leads to a contradiction because

$$0 = \lim_{k\to\infty} \frac{\ln(k\theta'(k))}{k} = \lim_{k\to\infty} \frac{\ln(\theta'(k))}{k} = \lim_{k\to\infty} \frac{\theta''(k)}{\theta'(k)} < 0,$$

where the first equality follows from the fact that $\lim_{k\to\infty} k\theta'(k) > 0$ and is bounded, the second is straightforward, and the third equality holds by L'Hôpital's rule.

CASE 2: Now suppose that $\lim_{k\to\infty} \psi^*(k) = 1$. We first take the limit of (B.6) as $k \to \infty$ to obtain $\lim_{k\to\infty} k\theta''(k) = -\frac{2(\lambda+1)\beta}{\sigma^2} < 0$. Note that $\lim_{k\to\infty} \psi^*(k) = 1$ implies $\lim_{k\to\infty} k(a - e^*(k)) = 0$, which in turn implies $\lim_{k\to\infty} ke^{*\prime}(k) = 0$. With this in mind, we multiply both sides of (B.13) by $k\theta'(k)$ and take the limit. Then it can be easily shown that the terms (i), (iii), and (iv) in (B.13) converge to zero. For the limit value of (ii), we substitute the ODE (B.5) for $\theta'(k)$ and simplify into

$$\lim_{k\to\infty} \frac{e^{*\prime}(k)}{a - e^*(k)} \cdot k\theta'(k) = \lim_{k\to\infty} \frac{a\lambda(k+h)e^{*\prime}(k)}{\sigma^2\psi^*(k)\left[-\lambda a + (\lambda+1)e^*(k)\psi^*(k)\right]} = 0,$$

where the last equality results from $\lim_{k\to\infty}(k+h)e^{*\prime}(k) = 0$. Therefore, the equation (B.13) leads to $\lim_{k\to\infty} k\theta''(k) = 0$, which contradicts with the negative limit value obtained from

(B.6). The proof is complete. $\square$

LEMMA 5. *If* $\lim_{k\to\infty}\psi^*(k)$ *exists, then* $\lim_{k\to\infty}\theta'(k)=0$.

PROOF OF LEMMA 5: Recall that when $\psi^*(k)\to\frac{\lambda}{\lambda+1}$, the numerator of $\theta'(k)$ in the ODE (B.9) approaches a constant, but the denominator takes an indeterminate form because $f(k)\equiv k\left[-\lambda a+(\lambda+1)e^*(k)\psi^*(k)\right]\to\infty\cdot 0$. Nevertheless, below we demonstrate that $f(k)$ diverges to ∞, which drives $\theta'(k)$ approaching zero. In order to show this, we subdivide the set of possible values of $\lim_{k\to\infty}f(k)$ into three cases: (i) $\lim_{k\to\infty}f(k)=0$; (ii) $\lim_{k\to\infty}f(k)>0$ but bounded; and (iii) $\lim_{k\to\infty}f(k)=\infty$.[32] Similar to the preceding proof, we derive a contradiction for the first two cases, which establishes $f(k)\to\infty$ and $\theta'(k)\to 0$ as well.

CASE 1: Suppose $f(k)\to 0$. Then $\infty=\lim_{k\to\infty}\theta'(k)>A_1$, so the principal's value function would be larger than the value function in the first-best regime for a sufficiently large k, which is a contradiction.

CASE 2: Suppose that $f(k)$ approaches a positive constant. Then $\lim_{k\to\infty}\theta'(k)$ exists and takes on a positive value. We first show that the limit of $\theta''(k)$ does not exist. For this purpose, we multiply both sides of (B.10) by $G_e(e^*(k),k)$ and write it as

$$k^2 e^{*\prime}(k)\cdot\frac{G_e(e^*(k),k)}{k^2}=\theta''(k)+G_k(e^*(k),k).$$

Suppose that the limit of $k^2e^{*\prime}(k)$ exists. Then by L'Hôpital's rule, its limit value must be $\frac{1}{r(\lambda+1)}$.[33] When $f(k)$ approaches a positive constant, however, $\frac{G_e(e^*(k),k)}{k^2}$ converges to a negative one, meaning that $\lim_{k\to\infty}\theta''(k)$ approaches a negative number. This implies that $\theta'(k)$ converges to a negative constant, contradicting with $\theta'(k)>0$ for every $k>0$. Therefore, neither $\lim_{k\to\infty}k^2e^{*\prime}(k)$ nor $\lim_{k\to\infty}\theta''(k)$ exists in this case.

Divide both sides of (B.6) by $k\theta'(k)$, getting

$$\begin{aligned}\frac{(1+\lambda)\beta}{k\theta'(k)}&=\frac{1}{k}\left[\frac{1}{R}-he^*(k)-\left(\frac{1}{r}+\frac{1}{R}\right)\ln\theta'(k)-\frac{1}{r}\ln\frac{\psi^*(k)}{\lambda}\right]\\&\quad+\frac{\sigma^2\theta'(k)}{2}\left[1-2\frac{\psi^*(k)e^*(k)}{a}+\frac{\lambda+1}{\lambda}\left(\frac{\psi^*(k)e^*(k)}{a}\right)^2\right]\\&\quad+\left[\delta-e^*(k)+\frac{k+h}{2ak}e^*(k)^2+\frac{\theta(k)}{Rk}\right]-\frac{\sigma^2}{2}\frac{\theta''(k)}{\theta'(k)}.\end{aligned}\tag{B.14}$$

We now use (B.14) to derive a contradiction. Note that when $\psi^*(k)\to\frac{\lambda}{\lambda+1}$, $k\theta'(k)$ diverges to

[32] Because $e^*(k)<a$ and $\psi^*(k)\in\left[\frac{\lambda}{\lambda+1},1\right]$ for all $k>0$, the expression $-\lambda a+(\lambda+1)e^*(k)\psi^*(k)$ is always nonnegative. Hence $f(k)\geq 0$ for all $k>0$.

[33] Note that $\psi^*(k)\to\frac{\lambda}{\lambda+1}$ implies $\lim_{k\to\infty}(k+h)(a-e^*(k))=\lim_{k\to\infty}\frac{a-e^*(k)}{1/k}=\lim_{k\to\infty}k^2e^{*\prime}(k)=\frac{1}{r(\lambda+1)}$.

∞ by (B.9). Then the left-hand side of (B.14) tends to zero, but the right-hand side takes an indeterminate form due to the presence of $\theta''(k)/\theta'(k)$. For this reason, the function $f(k)$ does not approach any positive value, implying that $f(k)$ must diverge to ∞. $\square$

One immediate consequence of the preceding lemma is the following result:

COROLLARY 2. *If* $\lim_{k\to\infty} \psi^*(k)$ *exists and* $k^2 e^{*\prime}(k)$ *is bounded for all* $k > 0$, *then*

$$\lim_{k\to\infty} \frac{\psi^{*\prime}(k)}{\psi^*(k)} = \lim_{k\to\infty} \frac{\theta''(k)}{\theta'(k)} = 0.$$

PROOF OF COROLLARY 2: As we demonstrated in the proof of Lemma 4, $\lim_{k\to\infty} \psi^*(k) = \frac{\lambda}{\lambda+1}$ and $\lim_{k\to\infty} ke^{*\prime}(k) = 0$ under the assumption of bounded $k^2 e^{*\prime}(k)$. This in turn implies that the derivative of $\psi^*(k)$ in (B.12) tends to zero as $k \to \infty$. For $\theta''(k)/\theta'(k)$, it is a routine task to check that the first three terms in (B.13) would vanish. To figure out the limit of the term (iv), we multiply the top and bottom by k and then take the limit, getting

$$\lim_{k\to\infty} (\lambda+1) \frac{\psi^{*\prime} k e^*(k) + \psi^* e^{*\prime}(k) k}{k\left[-\lambda a + (\lambda+1)\psi^* e^*(k)\right]} = \lim_{k\to\infty} (\lambda+1) \frac{\psi^{*\prime} k e^*(k)}{k\left[-\lambda a + (\lambda+1)\psi^* e^*(k)\right]} = 0.$$

Recall that the denominator now diverges to ∞ as we have proved in Lemma 5. The first equality follows from $e^{*\prime}(k)k \to 0$ and the second follows from the fact that $\psi^{*\prime}(k)k$ is bounded. $\square$

Proof of Proposition 5

We first compute pay-performance sensitivity in the optimal contract when firm size is large enough. For this, we exploit the above lemmas to show that every term in (4.3), except the first, vanishes in the limit as $k \to \infty$. Hence $\gamma_c(k)$ converges to $a/2$.

We begin with the last term in (4.3), which consists of 3 terms in the bracket. The first two terms in the bracket converge to zero by Corollary 2, and the last also converges to zero by Lemma 5 and the fact that both $e^*(k)$ and $\psi^*(k)$ are bounded. The second term in (4.3) vanishes as well, because $(k+h)e^{*\prime}(k) \to 0$ results from the assumption of $k^2 e^{*\prime}(k)$ being bounded and $e^{*\prime}(k) \to 0$ by Lemma 2. Therefore, we obtain $\lim_{k\to\infty} \gamma_c(k) = a/2$.

In case $k \to 0$, direct computation yields

$$\lim_{k\to 0} \gamma_c(k) = \frac{a}{2} + he^{*\prime}(0) - \frac{1}{r}\left\{\frac{\theta''(0)}{\theta'(0)} + \psi_k^*(0) + \psi_e^*(0)e^{*\prime}(0) - \frac{1}{\lambda}\theta'(0)\right\}$$

$$= \frac{a}{2} + \frac{(\theta'(0))^2 - \lambda\theta''(0)}{\theta'(0)R},$$

where we used $e^*(0) = a$, $\psi^*(0) = 1$, and $\psi_e^*(0) = rh$. To prove that $\gamma_c(0)$ is larger than $a/2$, recall $(\theta'(k))^2 > (\lambda+1)\theta''(k)$ for all $k \in [0,\infty)$ which is a necessary condition for the principal's value function to be concave. Hence $\frac{(\theta'(0))^2 - \lambda\theta''(0)}{\theta'(0)R}$ must be positive. □

Proof of Proposition 6

Using the expression of $I^F(k)$ in Appendix A and the expression of $I^*(k)$ in Appendix B, we simplify their difference into

$$\begin{aligned} I^F(k) - I^*(k) = &\frac{k+h}{2a}(a - e^*(k))^2 + \frac{1}{R}[\theta(k) - A_1 k] - \left(\frac{1}{r} + \frac{1}{R}\right)\ln\theta'(k) \\ &- \frac{ah}{2} - \frac{1}{r}\ln\frac{\psi^*(k)}{\lambda} - (1+\lambda)\left(\frac{\beta}{A_1} - \frac{1}{R}\right). \end{aligned} \tag{B.15}$$

Recall that according to the function $\phi(x)$ defined in Lemma 3, the two values $\theta(0)$ and $\theta'(0)$ are interrelated in the following manner:

$$\theta(0) = \frac{Rha}{2} - \lambda\ln\lambda + (1+\lambda)\left[\frac{R\beta}{\theta'(0)} - 1 + \ln\theta'(0)\right].$$

Using this relationship and the fact that $\lim_{k\to0}\psi^*(k) = 1$ and $\lim_{k\to0} e^*(k) = a$, the limit of the difference in (B.15) as $k \to 0$ reduces to

$$\lim_{k\to0}\left(I^F(k) - I^*(k)\right) = \beta(1+\lambda)\left(\frac{1}{\theta'(0)} - \frac{1}{A_1}\right).$$

Because $\theta'(0) \le A_1$ as we proved in Lemma 3, there is under-investment when firm size is sufficiently small.

To see over-investment at the other extreme $k \to \infty$, notice that every term on the second line, on top of the very first term on the first line in (B.15), is bounded as we have verified above. The remaining terms $\frac{1}{R}[\theta(k) - A_1 k] - \left(\frac{1}{r} + \frac{1}{R}\right)\ln\theta'(k)$ diverge to $-\infty$, because $\lim_{k\to\infty}\theta'(k) = 0$ by Lemma 5. □

References

Ai, H. and Li, R. (2015). Investment and ceo compensation under limited commitment. *Journal of Financial Economics*, **116** (3), 452–472.

ALBUQUERQUE, R. and HOPENHAYN, H. A. (2004). Optimal Lending Contracts and Firm Dynamics. *Review of Economic Studies*, **71** (2), 285–315.

— and WANG, N. (2008). Agency Conflicts, Investment, and Asset Pricing. *Journal of Finance*, **63** (1), 1–40.

BAKER, G. P. and HALL, B. J. (2004). CEO Incentives and Firm Size. *Journal of Labor Economics*, **22** (4), 767–798.

BIAIS, B., MARIOTTI, T., ROCHET, J.-C. and VILLENEUVE, S. (2010). Large Risks, Limited Liability, and Dynamic Moral Hazard. *Econometrica*, **78** (1), 73–118.

BOTTAZZI, G., COAD, A., JACOBY, N. and SECCHI, A. (2011). Corporate growth and industrial dynamics: evidence from French manufacturing. *Applied Economics*, **43** (1), 103–116.

CLEMENTI, G. L., COOLEY, T. and GIANNATALE, S. D. (2010). A Theory of Firm Decline. *Review of Economic Dynamics*, **13** (4), 861–885.

— and HOPENHAYN, H. A. (2006). A Theory of Financing Constraints and Firm Dynamics. *The Quarterly Journal of Economics*, **121** (1), 229–265.

COOLEY, T. F. and QUADRINI, V. (2001). Financial Markets and Firm Dynamics. *American Economic Review*, **91** (5), 1286–1310.

COX, J. C., INGERSOLL, J., JONATHAN E and ROSS, S. A. (1985). A Theory of the Term Structure of Interest Rates. *Econometrica*, **53** (2), 385–407.

DEMARZO, P. M. and FISHMAN, M. J. (2007). Optimal Long-Term Financial Contracting. *Review of Financial Studies*, **20** (6), 2079–2128.

—, —, HE, Z. and WANG, N. (2012). Dynamic Agency and the q Theory of Investment. *Journal of Finance*, **67** (6), 2295–2340.

— and SANNIKOV, Y. (2006). Optimal Security Design and Dynamic Capital Structure in a Continuous-Time Agency Model. *Journal of Finance*, **61** (6), 2681–2724.

DOW, J., GORTON, G. and KRISHNAMURTHY, A. (2005). Equilibrium Investment and Asset Prices under Imperfect Corporate Control. *American Economic Review*, **95** (3), 659–681.

EDMANS, A., GABAIX, X. and LANDIER, A. (2009). A Multiplicative Model of Optimal CEO Incentives in Market Equilibrium. *Review of Financial Studies*, **22** (12), 4881–4917.

EVANS, D. S. (1987). Tests of Alternative Theories of Firm Growth. *Journal of Political Economy*, **95** (4), 657–74.

FUDENBERG, D., HOLMSTROM, B. and MILGROM, P. (1990). Short-term contracts and long-term agency relationships. *Journal of Economic Theory*, **51** (1), 1–31.

GABAIX, X. and LANDIER, A. (2008). Why Has CEO Pay Increased So Much? *The Quarterly Journal of Economics*, **123** (1), 49–100.

GIBBONS, R. and MURPHY, K. J. (1992). Optimal incentive contracts in the presence of career concerns: Theory and evidence. *Journal of Political Economy*, **100** (3), 468–505.

GREENWOOD, J., HERCOWITZ, Z. and KRUSELL, P. (1997). Long-Run Implications of Investment-Specific Technological Change. *American Economic Review*, **87** (3), 342–62.

GRYGLEWICZ, S. and HARTMAN-GLASER, B. (2015). *Dynamic Agency and Real Options*. Tech. rep., Erasmus University Rotterdam and University of California at Los Angeles.

HALL, B. H. (1987). The Relationship between Firm Size and Firm Growth in the U.S. Manufacturing Sector. *Journal of Industrial Economics*, **35** (4), 583–606.

HE, Z. (2009). Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion. *Review of Financial Studies*, **22** (2), 859–892.

— (2011). A model of dynamic compensation and capital structure. *Journal of Financial Economics*, **100** (2), 351–366.

HOLMSTRÖM, B. (1979). Moral hazard and observability. *Bell Journal of Economics*, **10** (1), 74–91.

HOLMSTRÖM, B. and MILGROM, P. (1987). Aggregation and linearity in the provision of intertemporal incentives. *Econometrica*, **55** (2), 303–28.

HYMER, S. and PASHIGIAN, P. (1962). Firm Size and Rate of Growth. *Journal of Political Economy*, **70**, 556.

JENSEN, M. C. (1986). Agency Costs of Free Cash Flow, Corporate Finance, and Takeovers. *American Economic Review*, **76** (2), 323–29.

— and MURPHY, K. J. (1990). Performance Pay and Top-Management Incentives. *Journal of Political Economy*, **98** (2), 225–64.

JONES, L. E. and MANUELLI, R. E. (2005). Neoclassical Models of Endogenous Growth: The Effects of Fiscal Policy, Innovation and Fluctuations. In P. Aghion and S. Durlauf (eds.), *Handbook of Economic Growth*, *Handbook of Economic Growth*, vol. 1, *1*, Elsevier, pp. 13–65.

MURPHY, K. J. (1999). Executive compensation. In O. Ashenfelter and D. Card (eds.), *Handbook of Labor Economics*, *Handbook of Labor Economics*, vol. 3, *38*, Elsevier, pp. 2485–2563.

RADNER, R. (1985). Repeated Principal-Agent Games with Discounting. *Econometrica*, **53** (5), 1173–98.

ROGERSON, W. P. (1985). Repeated moral hazard. *Econometrica*, **53** (1), 69–76.

SANNIKOV, Y. (2008). A continuous-time version of the principal-agent problem. *Review of Economic Studies*, **75** (3), 957–984.

SCHAEFER, S. (1998). The Dependence Of Pay–Performance Sensitivity On The Size Of The Firm. *The Review of Economics and Statistics*, **80** (3), 436–443.

SCHÄTTLER, H. and SUNG, J. (1993). The First-Order Approach to the Continuous-Time Principal-Agent Problem with Exponential Utility. *Journal of Economic Theory*, **61** (2), 331–371.

SPEAR, S. E. and SRIVASTAVA, S. (1987). On repeated moral hazard with discounting. *Review of Economic Studies*, **54** (4), 599–617.

STEIN, J. C. (2003). Agency, information and corporate investment. In G. Constantinides, M. Harris and R. M. Stulz (eds.), *Handbook of the Economics of Finance*, *Handbook of the Economics of Finance*, vol. 1, *2*, Elsevier, pp. 111–165.

WILLIAMS, N. (2013). *On Dynamic Principal-Agent Problems in Continuous Time*. Working paper, University of Wisconsin-Madison.

YAMADA, T. and WATANABE, S. (1971). On the uniqueness of solutions of stochastic differential equations. *Journal of Mathematics of Kyoto University*, **11** (1), 155–167.

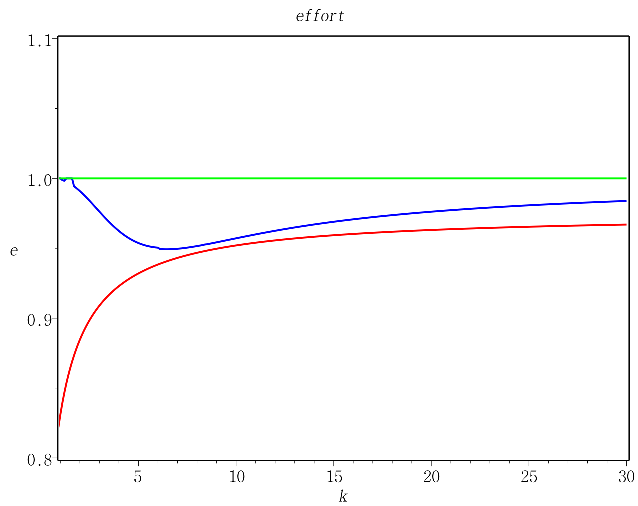

Figure 1: The Optimal Effort Policy over Firm Size for $h = a = 1$, $\beta = \delta = 0.01$, $R = 0.05$, $r = 4$, and $\sigma = 0.28$. The straight line displays the first-best effort policy $e^F(k) = a$, the upper curve the second-best one $e^*(k)$, and the lower curve the lower bound $e^\dagger(k)$.

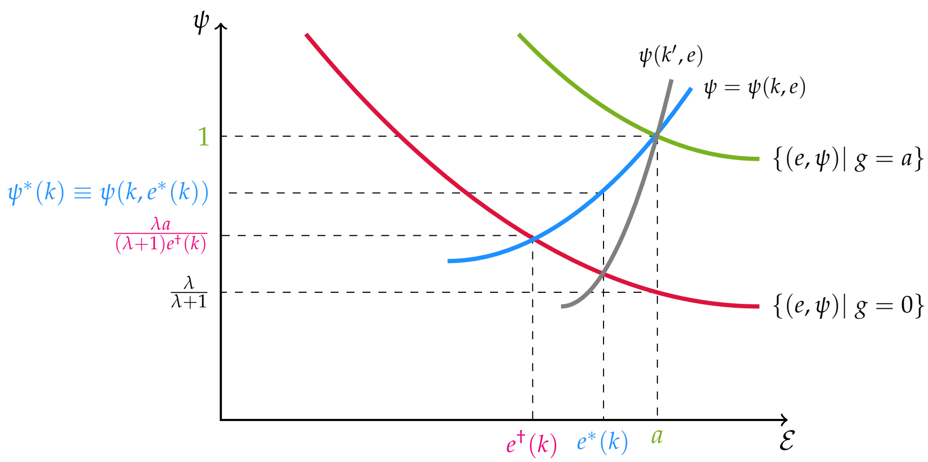

Figure 2: The two curves, labeled by $\{(e, \psi)|\ g = 0\}$ and $\{(e, \psi)|\ g = a\}$, are a level set of the function g where g takes on zero and $a > 0$, respectively. The other curve $\psi = \psi(k, e)$ describes the trajectory of the function ψ (regarding ψ as a function of e but holding k fixed) in the neighborhood of $e = a$. As k increases to k', the curve becomes steeper but always passes through the point $(a, 1)$. Note that by definition of $e^\dagger(k)$, the point $\left(e^\dagger(k), \frac{\lambda a}{(\lambda+1)e^\dagger(k)}\right)$ should be located at the intersection of $\{(e, \psi)|\ g = 0\}$ and ψ. Hence, the result $e^\dagger(k) < e^*(k) < a$ in part (a) leads to $\frac{\lambda}{\lambda+1} < \psi^*(k) < 1$.

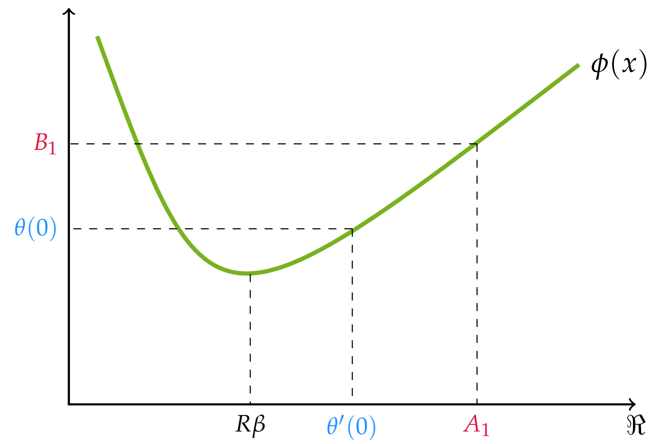

Figure 3: By Assumption 1 ($R\beta \leq A_1$), the function $\phi(x)$ must increase with x in a neighborhood of $x = A_1$. Because $\phi(\theta'(0)) = \theta(0) \leq B_1$ and $\theta'(0) > R\beta$ by Assumption 2, we have $\theta'(0) \leq A_1$.

	Pareto-Optimal Risk Sharing	**Second-Best Risk Sharing**
Compensation	$\frac{a}{2} + \frac{\lambda}{\lambda+1} \cdot \frac{A_1}{R}$	$\frac{a}{2}$
Dividend	$\frac{1}{\lambda+1} \cdot \frac{A_1}{R}$	0
Investment	$\frac{a}{2} - \frac{A_1}{R}$	$\frac{a}{2}$
Sum	a	a

Table 1: Comparison of Risk-Sharing Policies in a Large Firm - Each entry in the middle column indicates the Pareto-optimal sensitivity of a contractual term to changes in firm value k. The entry in the last column indicates the "limit value" of the sensitivity as $k \to \infty$ in the optimal contract.