A Quantitative Model of Dynamic Moral Hazard

Hengjie Ai, Dana Kiku, and Rui Li *

July 16, 2022

Abstract

We develop an equilibrium model with moral hazard, which arises because some productivity shocks are privately observed by firm managers only. We characterize the optimal contract and its implications for firm size, growth, and managerial pay-performance sensitivity and exploit them to quantify the severity of the moral hazard problem. Our estimation suggests that unobservable shocks are relatively modest and account for about 10% of the total variation of firm output. Nonetheless, moral-hazard induced incentive pay is quantitatively significant and accounts for 50% of managerial compensation. Eliminating moral hazard would result in about 1% increase in aggregate output.

Keywords: Dynamic Moral Hazard, CEO Compensation, Firm Dynamics

*Hengjie Ai (hengjie@umn.edu) is affiliated with the Carlson School of Management, University of Minnesota, Dana Kiku (dka@illinois.edu) is at the University of Illinois at Urbana-Champaign, and Rui Li (abclirui@gmail.com) is at the University of Massachusetts Boston. We would like to thank Dean Corbae, Steven Durlauf, Vivian Fang, Kenichi Fukushima, Zhiguo He, Antonio Mello, Adriano Rampini, Marzena Rostek, Vish Viswanathan, Noah Williams, Jun Yang, Amir Yaron, Tao Zha, seminar participants at the Fuqua School of Business at Duke University, Federal Reserve Bank of Atlanta, University of Calgary, University of Minnesota, and University of Wisconsin-Madison, and participants of the Tel Aviv Finance Conference, the ASU Sonoran Winter Conference, and the UBC Winter Finance Conference for useful comments on the topic. All errors are our own.

1 Introduction

A large body of literature in finance and economics emphasizes the importance of moral hazard in determining managerial compensation, firm investment, and firm growth. The vast majority of this literature focuses only on its *qualitative* implications. In this paper, we develop a general equilibrium model of dynamic moral hazard to analyze the *quantitative* impact of moral hazard on both individual firm behavior and the aggregate economy.

We incorporate moral hazard in a neoclassical production economy with heterogeneous firms. In our model, firm output is determined by both managers' investment decisions and productivity shocks. Because output is only a noisy signal of investment, managers have an incentive to substitute investment for private benefit. We allow firm productivity to be driven by both observable and unobservable shocks, and we use information from both time series and the cross section to identify their magnitudes in order to quantify the impact of moral hazard.

The key feature that distinguishes our model from the previous literature is the presence of both observable and unobservable productivity shocks. For theoretical convenience, most of the existing moral hazard models assume that all shocks are unobservable to the principal. However, observable shocks may account for a large fraction of the firm-level variation in productivity. Because the key tradeoff in moral hazard models is between risk sharing and incentive provision, the magnitude of observable shocks relative to unobservable shocks is the main determinant of the quantitative impact of moral hazard.

We derive two analytical results which allow us to use empirical evidence on the cross-sectional distribution and time-series dynamics of firms to identify the relative importance of observable and unobservable shocks. First, we define log-log pay performance sensitivity (PPS hereafter) to be the percentage change in CEO wealth for one percent increase in firm value and show that under the optimal contract, it is a decreasing function of managers' equity share in the firm. Intuitively, when risk-averse managers own a large fraction of firm's equity, diversification is poor and the welfare cost of incentive provision is high. As a result, the optimal contract specifies a low PPS.[1]

Second, we show that managers' share in the firm decreases after positive observable shocks but increases after positive unobservable shocks. Risk sharing requires that positive *observable* productivity shocks, which increase the total size of the firm, should lead to a reduction in the manager's share in the firm. The need for incentive provision implies that positive *unobservable* productivity shocks are followed by an increase in manager's share in the firm.

The above theoretical insights allow us to identify the magnitude of observable and unobservable shocks in the data using the size-investment relationship and the size-PPS relationship. In particular, if the size of unobservable shocks is relatively small, the need for risk sharing dominates, and managers' equity share in the firm must be decreasing in firm size. That is, as firm size declines, managers own a larger fraction of firm cash flow. Hence, their incentives for high investment and growth are largely aligned with shareholders' interests and marginal utility of stealing is low. Thus, when observable shocks are prevalent, investment is inversely related to firm size whereas pay-performance sensitivity is positively related to size. In contrast, when the magnitude of unobservable shocks is large, incentive provision requires managerial compensation to grow faster than firm size, which in equilibrium leads to a flat or even positive relationship between firm size and manager's equity share. Consequently, when moral hazard is severe, small firms invest at a lower rate and have a higher log-log PPS compared with large firms. Thus, we can exploit the cross-sectional variation in investment and PPS to infer the relative amount of private information.

Building on our theoretical results, we estimate the magnitude of observable and unobservable shocks using the Simulated Method of Moments (SMM) by exploiting moment

[1] A commonly used pay-performance sensitivity measure is the "level-level" PPS, which is the dollar increase in CEO wealth per dollar increase in firm value, which equals the product of the "log-log" PPS and manager's equity share. Consistent with the literature, our model implies that the level-level PPS is increasing in manager's equity share. We use the log-log PPS measure because the distribution of firm size in our model has a fat tail and the log-log PPS has the advantage of being scale invariant.

conditions of the joint distribution of firm size, investment and managerial compensation. We find that in order to account for the strong inverse relationship between firm growth and size and the positive relationship between pay-performance sensitivity and size observed in the data, the magnitude of unobservable shocks must be relatively small compared with observable shocks. Our estimates suggest that only about 10% of the total variation in firm output is attributed to unobservable shocks. The difference in variances of observable and unobservable shocks is statistically significant.

To provide further support for our model, we explore the implications of aggregate risk (i.e., beta) for the distribution of firm size, investment and CEO compensation. We show that due to optimal risk sharing, exposure to aggregate shocks raises investment rates and strengthens the negative relationship between manager's equity share and firm size. Consequently, our model predicts higher investment and a stronger PPS-size relationship among high-beta firms compared with low-beta firms. Intuitively, because aggregate news are observable, risk sharing of aggregate shocks strengthens the size-PPS relationship. In addition, because on average the aggregate economy is growing, high exposure to aggregate risks implies high expected growth and requires high investment. To ensure large enough investment to fuel future growth, high-beta firms have to provide managers with strong enough incentives. We show that these predictions are strongly supported by the data.

Based on the structural estimation of our model, we evaluate the quantitative impact of moral hazard in general equilibrium. In our dynamic model, moral hazard reduces efficiency for several reasons. First, incentive provision requires managerial compensation to respond to unobservable shocks and, therefore, reduces risk sharing. Second, moral hazard leads to back-loaded compensation policies and, hence, affects the intertemporal allocation of managerial compensation. Third, moral hazard distorts firms' investment policies and lowers the steady-state capital in the economy. Our estimates suggest that incentive provision is quite costly — incentive pay accounts for about 50% of the total CEO compensation. Our decomposition analysis reveals that about one third of incentive pay is due to limited risk sharing and the other two thirds are the compensation for distortions in the intertemporal allocation of CEO compensation. We also evaluate the aggregate impact of investment distortions and find that eliminating moral hazard would increase the total output of the economy by about 1%.

Finally, we use our model as a laboratory to study the impact of policy proposals that are often suggested as potential measures to curtail CEO pay inequality. As a first step, we confirm that while our estimation does not explicitly target the observed inequality in CEO compensation, our model matches well the Lorenz curve and the Gini coefficient in the data. We then examine the implications of two policies: the first one imposes a limit on the log-log PPS of compensation contracts, and the second one introduces a limit on the CEO-pay-to-worker compensation ratio. Our analysis reveals the importance of taking into account the endogenous response of compensation contracts to policy initiatives. We show that in both cases, policies aimed at curbing CEO pay inequality end up raising it in steady state due to the need of incentive provision. That is, failing to account for the endogenous response of private contracts may mask the unintended and unforeseen consequences of policy proposals, which may amplify rather than curtail the equilibrium CEO pay inequality.

Several novel features of our model are important for the purpose of quantifying the impact of moral hazard. First, we allow for observable shocks, which are the key determinant of the quantitative impact of moral hazard. Second, we adopt the constant relative risk aversion (CRRA) preferences. For tractability, most of the continuous-time contracting models assume risk neutrality or constant absolute risk aversion. The CRRA preference allows us to quantify the tradeoff between incentive provision and risk sharing. The optimal contract is fully determined up to an ordinary differential equation and can be efficiently solved for the purpose of estimation. Third, the dynamic general equilibrium setup allows us to exploit the implication of the optimal dynamic contract for the cross-sectional distribution of firm characteristics to identify the structural model parameters. In addition, general equilibrium is essential for understanding the welfare implications of moral hazard. A reduction in moral hazard is associated with higher levels of investment and capital accumulation. In a partial equilibrium, taking prices as given, firm profit is linear in capital (Hayashi (1982)). However, total output is decreasing return to scale with respect to capital at the aggregate level. Thus, in order to evaluate the impact of moral hazard across all firms, it is important to account for the decreasing return to scale in the aggregate production function, which requires a general equilibrium setup.

Related literature Our theoretical framework builds on the literature on optimal dynamic contracting, especially continuous-time models. The continuous-time methodology allows for semi-closed form solutions and makes it possible for us to estimate the model. The optimal contracting problem in our set-up is related to Sannikov (2008), DeMarzo and Sannikov (2006), DeMarzo, Fishman, He, and Wang (2012), and Zhu (2013). To study the quantitative implications of moral hazard, different from the above literature, we consider a general equilibrium setup with neoclassical production technologies and risk averse preferences.

Within the continuous-time contracting literature, our paper is mostly related to models that link moral hazard to pay-to-performance sensitivity, for example, He (2009), Hoffmann and Pfeil (2010), Edmans, Gabaix, Sadzik, and Sannikov (2012), Li (2017), and Di Tella and Sannikov (2016), Hackbarth, Rivera, and Wong (2021), and papers that study the impact of moral hazard on firm dynamics, for example, Hartman-Glaser, Lustig, and Xiaolan (2019) and Chi and Jin (2017).[2] Related to our paper, Hartman-Glaser, Mayer, and Milbradt (2018) also emphasize the importance of risk sharing for pay-performance sensitivity in the context of moral hazard.

Several recent papers explore the quantitative impact of dynamic agency using structural estimation or calibration. Lustig, Syverson, and Van Nieuwerburgh (2011) study an optimal contracting model with firm entry and exit dynamics to provide an explanation for the

[2]Recent papers that study moral hazard in continuous-time setup include Hartman-Glaser, Piskorski, and Tchistyi (2012), Feng and Westerfield (2018), Feng (2018), Leung (2017), Williams (2011), Piskorski and Westerfield (2016), Biais, Mariotti, Plantin, and Rochet (2007), Gryglewicz, Mayer, and Morellec (2020).

increased importance of organizational capital and managerial inequality. Nikolov and Schmid (2016) study the quantitative implications of information frictions for firms' capital structure and investment policy. Nikolov and Whited (2014) focus on the relationship between agency conflicts and cash accumulation. Nikolov, Schmid, and Steri (2019) examine the determinants of corporate liquidity management. Xiaolan (2014) studies the quantitative impact of limited commitment on firm-worker risk sharing. Sun and Xiaolan (2019) structurally estimate a dynamic agency model of firm financing. Different from these papers, our paper focuses on the identification of the relative magnitude of observable and unobservable shocks and the quantitative impact of moral hazard on CEO compensation and aggregate output.

Our paper is also related to the broader literature on structural estimation in corporate finance. For example, Taylor (2010, 2013) estimates learning models of CEO pay and CEO turnover. Li, Whited, and Wu (2016) estimate a model with limited commitment to quantify the importance of collateral and taxes in firms' capital structure decisions. Li and Whited (2016) estimate an adverse selection model to study the pattern of capital reallocation over business cycles. Several recent papers of Margiotta and Miller (2000), and **?** (**?**, 2015) analyze the identification and welfare implications of moral hazard in models with history-independent contracts (Fudenberg, Holmstrom, and Milgrom (1990)). Different from the above papers, the endogenous dynamics of firm size and promised utility are the key to our analysis. The presence of observable shocks and the history dependence of the optimal contract allow our model to match the salient features of the joint distribution of firm size, growth and managerial compensation, and provide identification for the structural parameters of our model.

The rest of the paper is organized as follows. We describe the setup of our model in Section 2. We provide the solution to the optimal contracting problem and lay out our identification strategy in Section 3. Section 4 describes our structural estimation and presents quantitative results. Section 5 concludes.

2 Setup of the Model

In this section, we introduce our equilibrium model of investment and managerial compensation with moral hazard.

2.1 Preferences and Technology

Time is infinite and continuous. There is a continuum of firms in the economy indexed by j. A firm j combines capital stock $K_{j,t}$ and labor $N_{j,t}$ to produce output at time t according to the production function $Y_{j,t} = Z_t^{1-\alpha} K_{j,t}^{\alpha} N_{j,t}^{1-\alpha}$, where Z_t is aggregate productivity and $\alpha \in (0,1)$ is the capital share in the production technology. We assume that Z_t follows a geometric Brownian motion:

$$dZ_t = Z_t \left(\mu_Z dt + \sigma_Z dB_{Z,t} \right), \tag{1}$$

The aggregate shock $B_{Z,t}$ is observable, as it is fully revealed by aggregate quantities in equilibrium.

Suppressing time subscripts, the operating profit of project j is defined as

$$\pi \left(K_j \right) = \max_{N_j} \left\{ Z^{1-\alpha} K_j^{\alpha} N_j^{1-\alpha} - W N_j \right\}, \tag{2}$$

where W is the equilibrium wage in the spot labor market. We assume that the market for labor is perfectly competitive. In addition, the total labor supply of the economy is constant and normalized to 1. We show in Appendix A that a balanced growth path exists where the profit function is constant returns to scale (CRS) in capital: $\pi \left(K_j \right) = A K_j$, the equilibrium marginal product of capital A is constant over time, and both Z_t and W_t grow at the same rate.

Operating each project requires a manager, who is the only agent that has special skills in accumulating project-specific capital. The law of motion of capital is given by:

$$dK_{j,t} = K_{j,t}\left[\left(i_{j,t} - \delta\right)dt + \sigma^{T} dB_{j,t} + \mu_Z dt + \sigma_Z dB_{Z,t}\right], \tag{3}$$

where $i_{j,t} = \frac{I_{j,t}}{K_{j,t}}$ is the investment-to-capital ratio, δ is the depreciation rate, $dB_{j,t}$ is a vector of firm-specific productivity, and σ is a vector of the corresponding volatility parameters.

In our setup, capital $K_{j,t}$ is publicly observable but investment decisions, $I_{j,t}$, are known only to the manager. Firm owners cannot infer the actual amount of investment from the observable capital stock because a part of the Brownian motion shock is assumed to be unobservable. In particular, we assume that the vector of firm-specific shocks is a given by:

$$\sigma^{T} dB_j = \sigma_U dB_{U,j} + \sigma_O dB_{O,j}, \tag{4}$$

where the Brownian motion $B_{U,j}$ is unobservable to all but the manager who operates the project. The Brownian motion $B_{O,j}$ is public information. To keep our setup simple, here we assume that volatility is independent of firm size.[3]

At any time t, firm owners face the following budget constraint:

$$D_{j,t} + C_{j,t} + h\left(\frac{I_{j,t}}{K_{j,t}}\right) K_{j,t} = AK_{j,t}, \tag{5}$$

where $D_{j,t}$ is the amount of dividends, $C_{j,t}$ is managerial consumption, and $h\left(\frac{I_{j,t}}{K_{j,t}}\right) K_{j,t}$ is the total cost of investment. We use a standard quadratic adjustment cost function, $h(i) = i + \phi i^2$, where $\phi > 0$ is a technology parameter. We assume that consumption and investment decisions are privately observable by managers. Moral hazard arises because managers can always substitute investment for their own private consumption. The degree of moral hazard is determined by the relative magnitude of unobservable versus observable

[3]Empirically, volatility is known to decrease with firm size. Luttmer (2011) and **?** propose models in which volatility depends on size due to diversification or production network mechanisms.

shocks, $\frac{\sigma_U^2}{\sigma_O^2}$.

Because the production technology exhibits constant returns to scale, the capital accumulation technology is linear, and the noise in the capital accumulation technology is proportional to firm size, the specification of moral hazard in our model is linearly homogeneous. Our formulation is consistent with the "multiplicative" specification of the moral hazard problems advocated by Edmans, Gabaix, and Landier (2009), and Edmans and Gabaix (2016), and it allows us to capture the presence of moral hazard in firms with vast differences in size.[4]

We assume that managers are risk averse and maximize expected utility with constant relative risk aversion. The time-t continuation utility of the manager is given by:

$$U_t = \left\{ E_t \left[\int_0^\infty e^{-\rho s} C_{t+s}^{1-\gamma} ds \right] \right\}^{\frac{1}{1-\gamma}}, \tag{6}$$

where ρ is the discount rate of the manager, and γ is the coefficient of relative risk aversion. For convenience, we have normalized utility so that it is homogeneous of degree one with respect to consumption.

Firm owners also have a CRRA preference over intertemporal consumption plans $\{C_{F,t}\}_{t=0}^\infty$ with a risk aversion γ_F and a discount rate ρ_F. For simplicity, we assume that managers and firm owners have the same risk aversion: $\gamma_F = \gamma$. The representative firm owner is endowed with one unit of labor and ownership of all firms' capital stock. Its objective is to maximize firm value given by

$$E_0 \left[\int_0^\infty \Lambda_t D_t dt \right], \tag{7}$$

where Λ_t is the stochastic discount factor, which in equilibrium must be consistent with the marginal rate of substitution of firm owners.[5]

[4]The Edmans, Gabaix, and Landier (2009) model is a static model with risk neutrality. Dynamics and risk aversion are the key elements of our model that allow for identification of the model parameters that we carry out in Section 4.

[5]As we explain below, managers and firms exit the economy at Poisson distributed times. In the utility and present value calculations such as in Equations (6) and (7), consumption and cash flow are zero after

In our baseline specification, all firms are assumed to have the same exposure to aggregate productivity shocks. As we show in Appendix B.1, this assumption together with the condition $\gamma_F = \gamma$ imply that the presence of aggregate shocks does not contribute to the cross-sectional heterogeneity in firm investment and CEO compensation policies. We will relax this assumption in Section 4.5 to study the implications of aggregate risk for firm decisions.

2.2 Firm Entry and Exit

A measure κ of new firms and managers arrive in the economy per unit of time. Upon entry, these firms and managers immediately match with each other and start production under a mutually agreed compensation contract for the manager. We normalize the initial size of firm capital to 1 and we assume that the initial utility of the manager is determined by a parameter U_0, which reflects the relative bargaining power of managers and firms. At the same time, existing firms along with their managers continuously exit the economy at a Poisson rate κ. Firms and managers stop receiving any cash flow or consumption after their exit. The assumption on the entry and exit dynamics implies that in a stationary equilibrium, where entry equals exit, the total measure of firms in the economy is 1.

2.3 Profit Maximization

In our setup, high investment accelerates the accumulation of capital and increases output. However, because investment in capital is not observable and managers have incentives to substitute investment for consumption, shareholders' investment plan can be implemented only if managers find it optimal to follow. To induce investment and ensure that shareholders' and managers' interests are aligned, firm owners reward high output and punish low output. In a dynamic setting, these incentives are provided by conditioning future managerial

the exit.

compensation on past performance.

Below, we formally describe the optimal contracting problem. A contract is a sequence of dividends, managerial compensation, and investment policies, $\left(\left\{D_{j,t}\left(K_j^t, B_{O,j}^t\right)\right\}, \left\{C_{j,t}\left(K_j^t, B_{O,j}^t\right)\right\}, \left\{i_{j,t}\left(K_j^t, B_{O,j}^t\right)\right\}\right)_{t=0}^{\infty}$, that depends on the history of the realization of observables, which we denote by $K_j^t = \{K_{j,t}\}_{s=0}^t$, $B_{O,j}^t = \{B_{O,j,s}\}_{s=0}^t$. To save notations, we suppress firm subscript and write a contract as $\{D_t, C_t, i_t\}_{t=0}^{\infty}$. We assume that upon entry, managers sign a contract with firm owners that promises to deliver the initial life-time utility of $U_0 = u_0 K_0$. That is, under the consumption policy specified by the contract,

$$\left\{E_t\left[\int_0^{\infty} e^{-(\rho+\kappa)s}(\rho+\kappa)\, C_{t+s}^{1-\gamma} ds\right]\right\}^{\frac{1}{1-\gamma}} \geq u_0 K_0. \tag{8}$$

Recall that we normalize the initial level of capital to $K_0 = 1$. As a result, the parameter u_0 selects an allocation on the constrained efficient Pareto frontier for managers and firm owners.

Given a contract $\{D_t, C_t, i_t\}_{t=0}^{\infty}$, if the manager follows the dividend payout policy, $\{D_t\}_{t=0}^{\infty}$, but chooses an alternative investment policy, $\{\tilde{\imath}_t\}_{t=0}^{\infty}$, his continuation utility at time t can be written as:

$$U_t\left(\{\tilde{\imath}_s\}_{s=0}^{\infty}\right) = \left\{E_t\left[\int_0^{\infty} e^{-(\rho+\kappa)s}(\rho+\kappa)\left(C_{t+s} + h\left(i_{t+s}\right)K_{t+s} - h\left(\tilde{\imath}_{t+s}\right)K_{t+s}\right)^{1-\gamma} ds\right]\right\}^{\frac{1}{1-\gamma}}. \tag{9}$$

That is, at time $t+s$, if the manager chooses a lower investment rate ($\tilde{\imath}_{t+s}$) than what is specified under the contract (i_{t+s}), he can privately save $h\left(i_{t+s}\right)K_{t+s} - h\left(\tilde{\imath}_{t+s}\right)K_{t+s}$ units of capital and use it for consumption without being detected by the shareholders. Hence, a contract $\{D_t, C_t, i_t\}_{t=0}^{\infty}$ is incentive compatible if the investment policy specified by the contract is optimal from the manager's perspective, that is, if:

$$U_t\left(\{i_s\}_{s=0}^{\infty}\right) \geq U_t\left(\{\tilde{\imath}_s\}_{s=0}^{\infty}\right), \quad \text{for all } t, \tag{10}$$

and for all investment policies $\{\tilde{i}_s\}_{s=0}^{\infty}$.

We assume double-sided limited commitment of financial contracts, as in Ai, Kiku, Li, and Tong (2021).[6] We assume that upon default, managers can take away a fraction of the firm's assets but are forever excluded from the credit market.[7] That is, managers are not allowed to enter into any intertemporal risk-sharing contracts after default. As we show in Appendix B, the utility of the manager upon default is a linear function of capital: $u_{MIN}K_t$. Hence, limited commitment on the manager side requires

$$U_t\left(\{i_s\}_{s=0}^{\infty}\right) \geq u_{MIN}K_t, \quad \text{for all } t \geq 0. \tag{11}$$

Any compensation plan that violates condition (11) may lead to the manager defaulting on the contract. In the quantitative analysis in Section 4, we treat u_{MIN} as a parameter of the model and estimate it from the data.

We also assume that shareholders cannot commit to negative net present value (NPV) projects. This constraint requires that the net present value of the firm's cash flow stays positive at all times:

$$E_t\left[\int_t^{\infty} \frac{\Lambda_s}{\Lambda_t} D_s ds\right] \geq 0, \quad \text{for all } t \geq 0. \tag{12}$$

Shareholders choose a contract $\{D_t, C_t, i_t\}_{t=0}^{\infty}$ that maximizes the present value of firm's cash flow subject to the budget constraint in Equation (5), the incentive compatibility constraint in Equation (10), and the limited commitment constraints in Equations (11) and (12).

[6]Some type of limited commitment is required to make the moral hazard problem non-trivial. In the absence of frictions, the principal can typically implement the efficient allocation arbitrarily closely if she is allowed to apply extremely severe punishment to the agent (see Mirrlees (1974)). In reality, managers typically have a variety of outside options and can always choose to leave a firm. We formally model outside options by limited commitment.

[7]Similar specifications are used in Albuquerque and Hopenhayn (2004), Kiyotaki and Moore (1997), and Kehoe and Levine (1993).

2.4 Equilibrium Definition

An equilibrium consists of time series of stochastic discount factors and wages, $\{\Lambda_t, W_t\}_{t=0}^{\infty}$, and a collection of optimal contracts, one for each firm, $\{D_{j,t}, C_{j,t}, i_{j,t}, N_{j,t}\}_{t=0}^{\infty}$, such that,

1. Given equilibrium prices, firms choose the optimal contract to maximize firm value subject to the constraints (10), (11), and (12).

2. The equilibrium prices are chosen such that the goods market and the labor market clear. That is, for all t,

$$C_{F,t} + \int \left[C_{j,t} + h(i_{j,t})K_{j,t}\right] dj = \int Y_{j,t} dj, \quad \int N_{j,t} dj = 1, \tag{13}$$

and the stochastic discount factor is consistent with shareholder consumption, $\Lambda_t = e^{-\rho_F t} \left(\frac{C_{F,t}}{C_{F,0}}\right)^{-\gamma}$.

2.5 Recursive Formulation

Following the standard approach in the dynamic contracting literature, we construct the solution to the optimal contracting problem recursively by using promised utility as a state variable. In our case, policy functions depend on two state variables (K, U), where K is the size of the firm and U is the continuation utility promised to the manager. We can think of the state variables, (K, U), as a summary of the firm's type. As in Atkeson and Lucas (1992), we construct the equilibrium allocation recursively. First, for firms of each type (K, U), we specify the flow rate of dividend payout, managerial compensation and investment-to-capital ratio using the policy functions $D(K, U)$, $C(K, U)$, $i(K, U)$. Next, we specify the law of motion of the state variables (suppressing time and firm subscripts):

$$\frac{dK}{K} = \left[i(K, U) - \delta\right] dt + \sigma_O dB_O + \sigma_U dB_U + \mu_Z dt + \sigma_Z dB_Z, \tag{14}$$

and

$$\frac{dU}{U} = \left[-\frac{\rho+\kappa}{1-\gamma}\left(\left(\frac{C(K,U)}{U}\right)^{1-\gamma}-1\right)+\frac{1}{2}\gamma\left(g_O^2\sigma_O^2 \ +g_U^2\sigma_U^2+g_Z^2\sigma_Z^2\right)\right]dt$$
$$+ g_O\sigma_O dB_O + \ g_U\sigma_U dB_U + g_Z\sigma_Z dB_Z. \tag{15}$$

Equation (15) follows the formulation in Sannikov (2008) except that we use a stochastic differential utility representation of the preference so that utility is measured in consumption units (see Equation (6)). Here, g_O is the elasticity of continuation utility with respect to observable idiosyncratic shocks, g_U is the elasticity with respect to unobservable idiosyncratic shocks, and g_Z is the elasticity with respect to aggregate shocks. Intuitively, the policy functions g_O, g_U and g_Z describe the rules of assigning continuation utilities based on the realizations of the Brownian shocks. At time t, for a given level of promised utility U_t, the principal allocates the manager's continuation utility over time and states by choosing an instantaneous consumption flow, $C(K_t, U_t)$, and elasticities of continuation utility with respect to Brownian motion shocks, $g_{U,t}$, $g_{O,t}$, and $g_{A,t}$.

Because the production technology is constant returns to scale and utility functions are homogeneous, the optimal contracting problem is homogeneous in the state variable K. In Appendix B, we further show that the homogeneity of the problem implies that the value function, $V(K,U)$, depends only on firm-specific state variables, K and U, and not on aggregate productivity Z. In addition, $V(K,U)$ satisfies

$$V(K,U) = v\left(\frac{U}{K}\right)K\,, \tag{16}$$

for some function v. We define $u = \frac{U}{K}$ as the normalized continuation utility and use homogeneity to write normalized consumption, investment, and dividend as functions of u:

$$c(u) = \frac{C(K,U)}{K};\ i(u) = \frac{I(K,U)}{K};\ d(u) = \frac{D(K,U)}{K}. \tag{17}$$

Intuitively, this result holds because aggregate risk in firms' cash flow and shareholders' consumption is identical. Because managers and shareholders have the same risk aversion, optimal risk sharing requires that consumption of both shareholders and managers has the same exposure to aggregate risk. As shown in Appendix B.1, the elasticities g_O and g_U are also functions of u, and $g_Z = 1$.

Since K is firm size and U is the value of the manager's future compensation package, we can intuitively interpret u as the manager's share in the firm. In what follows, we use the terminology of the normalized utility and the manager's equity share interchangeably.[8]

3 Optimal Contracting

3.1 Characterization of the Optimal Contract

In Lemma 1 in Appendix B, we provide the characterization of incentive compatibility. We show that an investment policy is incentive compatible if and only if the normalized policy function g_U satisfies

$$g_U(u) = (\rho + \kappa)\, c(u)^{-\gamma}\, u^{\gamma - 1} \cdot h'(i(u)) \tag{18}$$

for all u.

Recall from Equation (15) that g_U is the elasticity of managers' promised utility with respect to unexpected changes in firm cash flow due to unobservable shocks. That is, $g_U = \frac{\partial \ln U}{\partial \ln K}\Big|_{B_U}$ is the log-log measure of pay-performance sensitivity upon unobservable shocks.[9] To understand the incentive compatibility condition (18), note that optimality

[8]It is important to note that while manager's equity share and u are related, they are not identical. In fact, the optimal contract in general cannot be replicated by equity compensation alone. A more rigorous way to construct manager's equity share is to solve for an implementation of the managerial compensation contract using financial assets such as firm equity and a risk-free asset, and define equity share as the equity owned by managers as a fraction of total firm value. The so-defined manager's equity share is a monotone function of u but not exactly u. We use the term "equity share" with the understanding that u is quantitatively a good approximation of this more rigorous notion of the manager's equity share.

[9]Hereafter, we will use "elasticity" and "log-log PPS" interchangeably.

from mangers' point of view requires them to equate the marginal cost of saving one unit of investment to its marginal benefit:

$$\left.\frac{\partial U}{\partial K}\right|_{B_U} = h'(i)\,(\rho+\kappa)\left(\frac{C}{U}\right)^{-\gamma}. \tag{19}$$

Here $\left.\frac{\partial U}{\partial K}\right|_{B_U}$ is the amount of utility increase per unit of unexpected increase in K due to unobservable shocks, that is, the level-level measure of pay-performance sensitivity. Thus, from the manager's perspective, $\frac{\partial U}{\partial K}$ is the marginal benefit of investment. On the right-hand side of Equation (19), $h'(i)$ is the marginal cost of investment in consumption good units, $(\rho+\kappa)\left(\frac{C}{U}\right)^{-\gamma}$ is the marginal utility, and the product of the two is the marginal cost of investment measured in utility units.[10] Because the log-log PPS and the level-level PPS are related by:

$$\left.\frac{\partial \ln U}{\partial \ln K}\right|_{B_U} = \left.\frac{\partial U}{\partial K}\right|_{B_U} \times \frac{K}{U}, \tag{20}$$

it follows from Equation (19) that $g_U = (\rho+\kappa)h'(i)\left(\frac{C}{U}\right)^{-\gamma}\frac{K}{U}$, which gives Equation (18).

Condition (18) reduces the requirement of incentive compatibility to restrictions on the policy functions for consumption and investment, and the sensitivity of the continuation utility to unobservable shocks. This allows us to characterize the value function as the solution to an HJB equation, which we describe in the following proposition. To save notation, we define $r = \rho_F + \gamma\mu_Z - \frac{1}{2}\gamma(1+\gamma)\sigma_Z^2$ to be the effective discount rate for firm cash flow (see Equation (36) in Appendix A).

Proposition 1. *The normalized value function, $v(u)$, satisfies the following HJB differential*

[10] $C^{-\gamma}$ is the marginal utility under standard addatively separable specification of the utility function. Under our homogenous of degree one formulation, marginal utility is $(\rho+\kappa)\left(\frac{C}{U}\right)^{-\gamma}$.

equation

$$0 = \max_{\substack{c,i,g_O,\\ g_U=(\rho+\kappa)c^{-\gamma}u^{\gamma-1}h'(i)}} \left\{ \begin{array}{c} A - c - h(i) + v(u)\left(i - r - \kappa - \delta + \mu_Z - \gamma\sigma_Z^2\right) \\ +uv'(u)\left[\frac{\rho+\kappa}{1-\gamma}\left(1 - \left(\frac{c}{u}\right)^{1-\gamma}\right) - (i - \delta + \mu_Z) + \frac{1}{2}\gamma\left(g_U^2\sigma_U^2 + g_O^2\sigma_O^2 + \sigma_Z^2\right)\right] \\ +\frac{1}{2}u^2 v''(u)\left[(g_U - 1)^2\sigma_U^2 + (g_O - 1)^2\sigma_O^2\right] \end{array} \right\} \tag{21}$$

on the domain $[u_{MIN}, u_{MAX}]$*, with the following boundary conditions:*

$$\lim_{u \to u_{MIN}} v''(u) = \lim_{u \to u_{MAX}} v''(u) = \infty, \ \textit{and } v(u_{MAX}) = 0.$$

Proof. See Appendix B.4. □

In Figure 1, we plot the normalized value function, $v(u)$. Under the optimal contract, the normalized continuation utility of the agent, $u = \frac{U}{K}$, stays in the bounded interval, $[u_{MIN}, u_{MAX}]$. The limited commitment constraint in Equation (11) requires $u_t \geq u_{MIN}$ because any feasible contract must provide the manager with a continuation utility at least as high as his outside option, $u_{MIN}K$. As u increases, the value function, $v(u)$, declines because a higher fraction of future cash flows is promised to the manager. Limited commitment on the shareholder side that requires the NPV of the project to be non-negative at all times imposes an upper bound on u: u_{MAX} such that $v(u_{MAX}) = 0$ and $u_t \leq u_{MAX}$ for all t. As Figure 1 shows, the value function is concave on its domain.

Note that on most of its domain, the value function is monotonically decreasing in promised utility u — allocating a higher utility to the manager implies a lower value for the firm. However, for u close to u_{MIN}, the value function is increasing. This feature does not appear in models with limited commitment (Ai and Li (2015)) but is common in models with moral hazard, for example, DeMarzo and Sannikov (2006). In this region, the optimal contract is not renegotiation proof and it is possible to simultaneously increase utility of both the shareholder and the manager. This arrangement, although ex post inefficient, is ex ante optimal because it allows to implement a wider range of punishment that helps induce stronger incentives.

The key tradeoff in moral-hazard models is between risk sharing and incentive provision. In our setup, their relative importance depends on the amount of unobservable versus observable shocks, $\frac{\sigma_U^2}{\sigma_O^2}$. Below, we show how the dynamics of continuation utility and firms' investment policies depend on $\frac{\sigma_U^2}{\sigma_O^2}$. These theoretical results allow us to identify the relative importance of observable and unobservable shocks by exploiting the cross-sectional distribution of CEO compensation and growth that we carry out in Section 4.

3.2 Investment Policy

Under the optimal contract, the investment policy, $i\left(u_t\right) = \frac{I_t}{K_t}$, is an increasing function of managers' normalized utility. This feature is consistent with a decline in the cost of incentive provision as u increases. Figure 2 shows the optimal investment rate, $i\left(u\right)$ (top panel), the term $\left(\rho+\kappa\right) c\left(u\right)^{-\gamma} u^{\gamma-1}$ (middle panel), and the optimal log-log pay-performance sensitivity, $g_U\left(u\right)$ (bottom panel), as functions of the normalized utility u.

As explained above, the term $\left(\rho+\kappa\right) c\left(u\right)^{-\gamma} u^{\gamma-1}$ is the percentage increase in managers' utility upon stealing an additional fraction of firm cash flow, or simply log-log marginal utility. As such, it can be interpreted as the marginal cost of incentive provision. The higher the marginal utility, the more tempting it is for managers to steal firm cash flow. As the middle panel of Figure 2 shows, the log-log marginal utility is decreasing in manager's equity share in the firm — when managers own a larger fraction of firm cash flow, the marginal utility of stealing is lower.

The optimal choice of log-log PPS, g_U, must trade off the benefit of incentive provision against the cost of imperfect risk sharing, which in utility terms is measured by $\frac{1}{2}\gamma(g_U\sigma_U)^2$, as shown in Equation (15). When u is low, the marginal cost of incentive provision is high. Thus, high incentives are needed to induce investment (the bottom panel of Figure 2). However, despite a high g_U, the investment rate is relatively low (the top panel). As u increases, the log-log marginal utility falls due to concavity of the utility function. As a result, it is possible to induce a higher investment rate with lower incentives. Therefore, as u increases, the log-log PPS declines but the investment rate increases.

Limited commitment also contributes to the positive relationship between the investment rate and the manager's share. When u approaches u_{MAX}, the limited commitment constraint on the shareholder side is likely to bind. A binding limited commitment constraint is associated with poor risk sharing. Therefore, as u increases, it is optimal for firms to increase investment and improve risk sharing.

3.3 Dynamics of Continuation Utility

In this section, we discuss the optimal response of continuation utility with respect to aggregate, observable, and unobservable shocks. Using Equations (14) and (15), we can write the law of motion of u as

$$\frac{du}{u} = \mu_u(u)dt + [g_O(u) - 1]\,\sigma_O dB_O + [g_U(u) - 1]\,\sigma_U dB_U, \tag{22}$$

where the function $\mu_u(u)$ is given in Equation (56) in Appendix C.

Note that the presence of aggregate shock does not affect the diffusion of normalized utility u. As shown in Equation (3), the cash flow of all firms have the same exposure to the aggregate shock $dB_{Z,t}$. Because shareholders and managers have the same risk aversion and because aggregate shocks are observable, the optimal contract features $g_Z = 1$. Intuitively, firm cash flow, which is proportional to $K_{j,t}$, and manager consumption have the same exposure to $dB_{Z,t}$. After normalizing by $K_{j,t}$, managers' utility $u_{j,t} = \frac{U_{j,t}}{K_{j,t}}$ does not respond to the aggregate Brownian motion shock.

Next, we provide a sufficient condition under which the drift of normalized utility, $\mu_u(u)$, also does not depend on the magnitude of aggregate shock, σ_Z.

Corollary 1. *Consider two economies that differ only in the drift and volatility of aggregate productivity growth, which we denote by $\mu_{i,Z}$ and $\sigma_{i,Z}$, respectively, for $i = 1, 2$. Suppose that one of the following conditions holds:*

1. *$\mu_{i,Z}$ and $\sigma_{i,Z}$ satisfy*
$$\mu_{1,Z} - \frac{1}{2}\gamma\sigma_{1,Z}^2 = \mu_{2,Z} - \frac{1}{2}\gamma\sigma_{2,Z}^2,$$
2. *$\gamma = 1$, that is, both shareholders and managers have log preferences,*

then the dynamics of the normalized utility and firm value functions in economy 1 and economy 2 are identical.

Proof. The HJB equations are identical as long as one of the above conditions are satisfied. □

Conditions 1 and 2 in Corollary 1 are essentially sufficient conditions for the two economies to have the same wealth-to-consumption ratio regardless of the value of σ_Z. Under these conditions, the ratio of the marginal rate of substitution of shareholders and managers does not depend on the magnitude of aggregate shock σ_Z, and therefore, the drift of the normalized utility u does not either. Because normalized CEO compensation and investment are only functions of u, the corollary provides a benchmark under which the presence of aggregate shock is irrelevant for the identification of the key structural model parameters. It is important to note that Corollary 1 depends crucially on the assumption that all firms have the same exposure to aggregate shock $B_{Z,t}$. In Section 4.5, we relax this assumption and study the implications of heterogeneous exposure to aggregate shocks.

While in our baseline model, the optimal contract requires that shareholder and managers' consumption have the same exposure to the aggregate shock B_Z, the responses to firm-specific shocks B_O and B_U have to be quite different. To understand the response of continuation utility with respect to observable and unobservable idiosyncratic shocks, in the top panel of Figure 3, we plot the elasticities of promised utility U with respect to observable shocks (g_O, solid line) and unobservable shocks (g_U, dashed line) implied by the optimal contract.

Intuitively, because $g_O = \frac{d \ln U}{d \ln K}\Big|_{B_O}$ is the log-log pay-performance sensitivity with respect to observable shocks, the choice of g_O is determined by risk sharing considerations. We define $\Gamma(u) = \frac{u \cdot v''(u)}{v'(u)}$ as the induced risk aversion of the shareholder's value function, i.e., it is the Arrow-Pratt measure of relative risk aversion of the shareholder's value function with respect to continuation utility.[11] We show in Appendix B.2, Equation (49), that the optimal choice of g_O depends on the relative risk aversion of the manager and the induced risk aversion of the shareholder: $g_O = \frac{\Gamma(u)}{\Gamma(u)+\gamma}$.

Without agency frictions, $\Gamma(u) = 0$, because the shareholder is well diversified and therefore is risk neutral with respect to idiosyncratic shocks. However, in our model, whenever the value function is strictly decreasing, $\Gamma(u) > 0$.[12] As in DeMarzo and Sannikov (2006), agency frictions induce risk aversion in the shareholder's value function. In fact, on the boundary as $u \to u_{MAX}$, $\Gamma(u) \to \infty$. Here, a small movement of u due to an observable shock leads to the violation of the limited commitment constraint ($u \leq u_{MAX}$), and the shareholder's value function is locally infinitely risk averse. As a result, as $\Gamma(u) \to \infty$, $g_O = \frac{\Gamma(u)}{\Gamma(u)+\gamma} \to 1$ — observable shocks are entirely passed through to managers. The same holds at the left boundary $u = u_{MIN}$.

Infinite risk aversion at the boundaries affects the shape of the value function in the interior because in a dynamic model, a non-trivial elasticity of u with respect to B_O in the interior increases the probability of a binding limited commitment constraint in the future and is welfare reducing. In general, whenever $\Gamma(u) > 0$, $g_O \in (0, 1)$ — the optimal contract provides some insurance against idiosyncratic shocks, but the insurance is imperfect and

[11]In most of its domain, the shareholder's value function is decreasing, therefore our risk aversion measure does not need a negative sign.

[12]Because the value function is globally concave, in the region of $v'(u) < 0$, risk aversion should be defined as $-\frac{u \cdot v''(u)}{v'(u)}$ and is still strictly positive.

shocks are partially passed through to managerial compensation. In regions far away from both boundaries, $\Gamma(u)$ is close to zero, and so is g_O.[13]

The top panel of Figure 3 also shows the log-log PPS with respect to unobservable shocks, g_U. In our model, risk sharing itself requires a positive pay-performance sensitivity. Without the incentive compatibility constraint in Equation (18), for risk sharing purposes, it is optimal to set g_U at the unconstrained optimal level, i.e., $\frac{\Gamma(u)}{\Gamma(u)+\gamma}$. However, in general, such PPS is not enough to induce the desired level of investment. In fact, risk sharing requires g_O to be positive but typically less than 1. However, the optimal g_U is higher than 1 in almost the entire domain, as Figure 3 shows. In addition, as explained above, g_U is in general decreasing in u due to the declining marginal cost of incentive provision.

The strength of incentives can also be measured using the level-level PPS, that is, $\frac{dU}{dK}$. Clearly, the level-level PPS and the log-log PPS are related by

$$\frac{dU}{dK} = \frac{d\ln U}{d\ln K}\frac{U}{K}. \tag{23}$$

The bottom panel of Figure 3 shows that the level-level measures of PPS with respect to observable and unobservable shocks, $g_O(u) \cdot u$ and $g_O(u) \cdot u$, respectively, are both increasing in u. Intuitively, holding the log-log PPS constant, firms where managerial compensation is a large fraction of firm cash flow, or equivalently, firms with a high manager's equity share (u), have a high level-level PPS.

In our model, manager's equity share is decreasing in firm size. Therefore, the cross-sectional variation in $i(u)$, $g_O(u)$ and $g_U(u)$ is consistent with the empirical evidence that small firms grow faster, have a higher level-level PPS and a lower log-log PPS than large firms. Because the log-log PPS in our model determines the size-investment and size-

[13]In the non-renegotiation proof region of u, $v'(u) > 0$: committing to punishing the manager for bad performance is ex ante desirable but ex post inefficient because it simultaneously lowers the shareholder's value. In this region, $g_O > 1$. Intuitively, it is beneficial for the optimal contract to promise high cash flow to the manager in the future, as it benefits both the manager and the shareholder. However, promise keeping requires lowering the current payment in order to raise future payment, which is implemented by creating excessive risk exposure under the optimal contract, i.e., $g_O > 1$.

PPS relationship, and because the differences between $\frac{d\ln U}{d\ln K}\Big|_{B_O}$ and $\frac{d\ln U}{d\ln K}\Big|_{B_U}$ are key for identification, in the rest of the paper, we focus on the log-log measure of incentives (which we simply refer to as PPS).

3.4 Implications for Identification

As discussed above, under the optimal contract, investment and growth rates increase with the normalized utility u, while pay-performance sensitivity decreases with u. These implications could potentially help identify the structural model parameters. However, the continuation utility is not directly observable. In this section, we show that we can identify the relative amount of private information and estimate the model using the readily available data on firm size by exploiting the relationship between size and continuation utility, which is endogenously determined by the optimal contract. In particular, we show that moral hazard determines the equilibrium relationship between firm size and the unobservable continuation utility. If the magnitude of unobservable shocks is relatively small, firm size and continuation utility are negatively correlated, whereas large values of $\frac{\sigma_U^2}{\sigma_O^2}$ imply a zero or even positive relationship between firm size and u. Because our model implies a monotone relationship between investment rate and u, and a near-monotone relationship between PPS and u, we can exploit the joint empirical distribution of growth rates, PPS and firm size to identify the model parameters.

In our model (see Equation (22)), $g_O - 1$ is the elasticity of u with respect to observable shocks, and $g_U - 1$ is the elasticity of u with respect unobservable shocks. As shown in Section 3.3, in most of the domain, $g_O(u) \leq 1$, $g_U(u)$ is significantly higher than $g_O(u)$, and typically $g_U(u) \geq 1$. Consider first the case in which most of the shocks are observable, i.e., $\frac{\sigma_U^2}{\sigma_O^2}$ is close to 0. In this scenario, the relationship between firm size and the normalized continuation utility is negative because it is mostly driven by observable shocks and $g_O \leq 1$. Intuitively, while a positive shock increases both firm size K and continuation utility U, risk sharing requires U to increase at a lower rate than K. Hence, the manager's share u falls, and K and u are negatively correlated.

If the contribution of unobservable shocks increases, that is, if $\frac{\sigma_U^2}{\sigma_O^2}$ becomes large, the negative correlation between firm size and manager's share weakens because $g_U > g_O$. In fact, because $g_U \geq 1$ in most of its domain, the correlation between K and u might be even positive. As moral hazard becomes severe, the cost of incentive provision increases; therefore, as firms grow and become larger, they have to provide a higher fraction of their cash flow to managers. We illustrate the relationship between firm size and the normalized utility in Figure 4. We consider several cases for the break-down between observable and unobservable shocks: $\frac{\sigma_U^2}{\sigma_O^2} = 0.1$ (solid line), $\frac{\sigma_U^2}{\sigma_O^2} = 1$ (dashed line), and $\frac{\sigma_U^2}{\sigma_O^2} = 10$ (dotted line). As the figure shows, when most shocks are observable, u is monotonically decreasing in firm size. As the relative magnitude of unobservable shocks increases, the negative relationship between u and K becomes considerably weaker and eventually changes its sign.

Figure 5 shows how the equilibrium relationship between firm size and the normalized utility translates into the relationship between size and investment. For each of the three cases, we plot the investment rate $i(u)$ as a function of firm size. For low levels of $\frac{\sigma_U^2}{\sigma_O^2}$, our model features a strong inverse relationship between investment rate and firm size, which is due to a strong negative correlation between the normalized utility and size. As the magnitude of unobservable shocks increases, the negative relationship between investment and size disappears and ultimately reverses to positive.[14]

Figure 6 shows the cross-sectional distribution of the average elasticity of continuation utility to productivity shocks. We compute the average elasticity, ξ, as a weighted average of elasticities with respect to observable and unobservable shocks, $\xi = \sqrt{\frac{\sigma_U^2}{\sigma^2} g_U^2 + \frac{\sigma_O^2}{\sigma^2} g_O^2}$,

[14]Note that a link between firm size and investment is not unique to our model. For example, the inverse relationship between the two can also be generated by models that feature a decreasing returns to scale technology (DRS). To explore the robustness of our evidence to returns to scale, we have estimated a model specification that allows for DRS. We find that decreasing returns to scale is rejected in favor of constant returns to scale and our key empirical finding of a relatively small magnitude of unobservable shocks continues to hold in the decreasing returns to scale specification.

where $\sigma^2 = \sigma_U^2 + \sigma_O^2$ is the total variance. Although ξ is not observable, it determines pay-performance sensitivity (i.e., the elasticity of CEO pay to firm performance), which we can measure using the available data.[15] Figure 6 illustrates two important implications that help identify the relative magnitude of observable and unobservable shocks. First, as the amount of unobservable shocks increases, the overall elasticity of managerial pay to firm performance rises due to the higher cost of incentive provision. Second, a relatively modest amount of unobservable shocks (eg., $\frac{\sigma_U^2}{\sigma_O^2} = 0.1$) implies an increasing (almost monotone) relationship between ξ and firm size. As the contribution of unobservable shocks gets larger, the relationship between the elasticity of managerial pay and firm size gradually becomes negative. To summarize, Figure 6 shows that both the overall level and the cross-sectional variation in PPS provide important information about the relative magnitude of observable versus unobservable shocks.

Building on these theoretical insights, in the next section, we estimate the model by exploiting moment conditions of the joint distribution of firm size, investment and managerial compensation.

4 Quantitative Evidence

4.1 Data

In our quantitative analysis, we exploit the panel data of US non-financial, non-utility firms that come from the Center for Research in Securities Prices (CRSP), Compustat, and Standard & Poor's ExecuComp database. For each firm in our sample, we collect its

[15]The mapping between the average elasticity of utility with respect to shocks, ξ, and the elasticity of managerial compensation with respect to firm performance (that is, the log-log based measure of PPS) is monotone, but nonlinear. For example, for $\frac{\sigma_U^2}{\sigma_O^2} = 0$, the log-log PPS is zero in the interior of (u_{MIN}, u_{MAX}) due to risk sharing. However, $\xi > 0$ because continuation utility accounts for a possibility of a binding constraint in the future, which is associated with a positive response of managerial pay with respect to shocks. Estimating the magnitude of $\frac{\sigma_U^2}{\sigma_O^2}$ from the level of PPS is thus a quantitative issue, which we address formally in the next section.

size, investment, managerial compensation and wealth. We measure firm size by the sum of physical capital (property, plant and equipment) and intangible capital. Similarly, firm investment is measured by investment in physical and intangible capital.[16] We rely on the total executive compensation figures in ExecuComp database to measure the flow of CEO compensation that comprises salary, bonuses, the value of restricted stock granted, the value of options exercised, and long-term incentive payouts. In addition, we construct a measure of CEO wealth as a sum of salary, bonuses, the market value of restricted shares held by executives, the market value of shares and stock options owned, the net revenue from stock trading, dividends, long-term incentive payments, and the present value of future payoffs. The measurement of CEO wealth follows the approach of Aggarwal and Samwick (1999), Himmelberg and Hubbard (2000), and Clementi and Cooley (2009), and it is described comprehensively in Appendix E. All nominal quantities are converted to real using the consumer price index provided by the Bureau of Labor Statistics. The data are sampled on the annual frequency and cover the period from 1992 till 2019. Because in estimation we exploit the cross-sectional differences in the dynamics of CEO wealth, we limit our sample to firms with available executive compensation data.

4.2 Structural Estimation

We estimate the model parameters using the simulated method of moments (McFadden (1989), and Pakes and Pollard (1989)). Our model has fourteen parameters in total but our primary focus is on volatility parameters that govern the magnitude of observable and unobservable shocks, σ_O and σ_U, respectively. Following the discussion in Section 3, our identification strategy is to exploit the dynamics of firm growth and executive pay-

[16] Physical investment is measured by the change in physical capital adjusting for depreciation. Intangible investment is inferred from the stock of intangible capital assuming the amortization rate of 10%, which is in line with Peters and Taylor (2017) and the current IRS guidelines that require most intangible assets to be amortized within a 15-year period. Our empirical evidence is robust to using capital expenditure as a measure of physical investment and to using alternative amortization rates of intangible capital within a reasonable range of 5%–25%.

performance sensitivity. Because firms' growth and CEO compensation are determined jointly by moral hazard and other model parameters, we also estimate risk aversion (γ), the discount rate (ρ), the productivity parameter (A), the capital depreciation rate (δ), the adjustment cost parameter (ϕ), and the parameters that determine the outside option of managers and the initial normalized utility (u_{MIN} and u_0, respectively).[17] Thus, together with the volatility parameters, we estimate a subset of nine structural parameters that have a first-order effect on the joint distribution of firm growth and managerial compensation.

The remaining five parameters either do not play a critical role in determining the cross-sectional distribution of firms' characteristics or cannot be identified separately. We calibrate them consistently with the standards of the macroeconomics literature to match a different set of moments. We choose $\sigma_Z = 3\%$, and $\mu_Z = 1.5\%$ to match volatility and average growth of aggregate capital. Following Kydland and Prescott (1982), King and Rebelo (1999) and Rouwenhorst (1995), we set the capital share at 0.33, and calibrate the death rate to be 0.05 to match the average firms' exit rate in the data. We calibrate the discount rate of shareholders at $\rho_F = 0.03$, which, together with the assumed aggregate consumption dynamics and estimated risk aversion, imply an effective discount rate of firm cash flow of $r + \gamma\sigma_Z^2 = 4.3\%$ per year, matching the average return of the risk-free treasury bills and risky assets in the data. The values of the calibrated parameters are reported in Panel B of Table 1.

To estimate the structural parameters, let

$$\Theta = \{\sigma_O,\, \sigma_U,\, \gamma,\, \rho,\, A,\, \delta,\, \phi,\, \bar{u}_{MIN},\, \bar{u}_0\} \tag{24}$$

denote the vector of parameters to estimate, and let $\mathcal{M}_D$ and $\mathcal{M}_M(\Theta)$ denote the vectors of data-based and model-implied moments, respectively.[18] We estimate the model parameters

[17] It is more convenient to estimate the equilibrium marginal product of capital directly. We can always use the equilibrium relationship between the marginal product of capital and the primitive productivity parameter to back out the latter (see Appendix A).

[18] For computational convenience, in estimation we use the following parameterization: $\bar{u}_{MIN} \equiv$

by minimizing the following objective function:

$$\widehat{\Theta} = \underset{\Theta}{\operatorname{argmin}} \big[\mathcal{M}_D - \mathcal{M}_M(\Theta)\big]'\Omega\big[\mathcal{M}_D - \mathcal{M}_M(\Theta)\big], \tag{25}$$

where Ω is the weighting matrix. The model-based moments are computed via simulations. Specifically, we draw a panel of shocks, and for a given parameter configuration we solve the model numerically, discretize it and simulate a cross-section of firms. The simulated panel consists of 10,000 firms per year and the length of time series is set at 150 years. We discard the first 100 years of data and use the remaining 50 years of simulated data to calculate the vector of moments $\mathcal{M}_M(\Theta)$. Because our panel is fairly large, the simulated moments represent the population moments sufficiently well. We confirm that increasing either the length of the simulated sample or the size of the cross section has virtually no effect on the model-implied moments and the parameter estimates. We estimate the model parameters using the optimal weight matrix. Hence, the estimation is carried out in two stages: in the first stage, we obtain the initial estimates by weighting each moment condition by the inverse of the variance of the sampling distribution of the corresponding statistic, and in the second stage we use the inverse of the variance-covariance matrix of the moment conditions evaluated at the first-stage estimates. We use the Newey and West (1987) estimator of the spectral density matrix at frequency zero with a truncation lag of two.

Guided by the model's implications discussed in Section 3, we exploit moments that are informative about the degree of moral hazard and that help identify the structural parameters. These moments include the slope of the power law of the distribution of firm size, the average standard deviation of firm growth rates, and the median growth and investment rates, and the log-log pay-performance sensitivity for a cross section of size-sorted portfolios.

To calculate moments of the joint distribution of firm size, growth and managerial compensation, we construct five size-sorted portfolios using breakpoints that are equally-

u_{MIN}/u_{MAX}, and $\bar{u}_0 \equiv u_0/u_{MAX}$.

spaced in log size. Portfolios are re-balanced at the annual frequency, and for consistency, in the data and in the model, firm size is measured by capital stock. The full list of moments that we exploit in estimation is presented in Table 2. In all, we use 17 moments to estimate 9 model parameters.

Figure 7 provides a graphical illustration of the relationship between firm size and average growth, investment and pay-performance sensitivity. The solid line in each panel shows the point estimates, and the two dotted lines represent the two-standard error band. As is well known, small firms, on average, invest more and grow at a much higher rate relative to large firms. For example, in our sample, the difference in average growth rates between firms in the smallest and largest portfolios is 9.5% with a robust t-statistic of about 2.8. As the bottom panel of Figure 7 reveals, pay-performance sensitivity also depends on firm size — the log-log measure of PPS, estimated by regressing changes in log wealth on log returns controlling for firm and time fixed effects, is monotonically increasing in size. As we discuss below, the cross-sectional dispersion in PPS is also statistically significant.

Although the model parameters are identified jointly by the equilibrium conditions, below we provide an intuitive discussion of which moments help the most in the identification of different model parameters. Our model has six technology parameters. The key focus of our estimation is on the relative importance of unobservable shocks, i.e., $\frac{\sigma_U^2}{\sigma_O^2}$. As follows from the discussion in Section 3.4, the ratio is identified by the cross-sectional relationships between firm size and investment, and firm size and PPS. The level of volatilities is then pinned down by the average standard deviation of firm growth rates. Further, the average growth rate across firms is informative about the marginal product of capital, A. The difference between growth rates and investment rates is determined by the depreciate rate, and therefore, the two sets of moments help identify δ. The investment-size relationship is informative about the adjustment cost parameter ϕ. To understand why, note that the incentive compatibility condition in Equation (18) relates investment rate to promised utility. In particular, the curvature of the adjustment cost function determines the sensitivity of investment rates with respect to u, and therefore, with respect to firm size (as discussed above, u is a monotone function of firm size).

The other parameters are identified as follows. First, risk aversion of managers is identified by the average log-log pay performance sensitivity, because PPS under the optimal contract is determined by the trade-off between incentive provision and risk sharing. Second, the growth rate and investment rate of large firms are informative about managers' outside option u_{MIN} because, as explained in Section 3.4, large firms tend to converge to the region close to u_{MIN}. In addition, as shown in Luttmer (2007) and Ai, Kiku, Li, and Tong (2021), the power law of firm size distribution is determined by the growth rates of large firms. Hence, the slope of the power law further helps identify u_{MIN}. Third, the manager's discount rate ρ is identified by the cross-sectional variation in PPS and growth. Intuitively, the difference between the effective discount rate and managers' discount rate determines the intertemporal allocation of managerial pay and, therefore, the growth in promised utility u_t. A higher ρ implies that managers are less patient and that, over time, promised utility converges faster to u_{MIN}. This in turn implies that u_t of large firms is more likely to exceed the outside option u_{MIN} and renders the manager-side limited commitment constraint binding. Therefore, given the interest rate, the life-cycle dynamics of PPS and investment rates are informative about the manager discount rate ρ. Finally, the initial promised utility $u_0 = \frac{U_0}{K_0}$ is identified by the cross-sectional variation in firm growth rates and investment rates. In our model, firms start small and grow over time. Because investment rates and average growth rates are determined by normalized utility u_t, the initial condition u_0 affects the size-investment and size-growth relationships in the model.

4.3 Parameter Estimates and Implications

Panel A of Table 1 presents the SMM estimates of the model parameters. First, notice that the set of moment conditions that we exploit in estimation allows us to identify the structural parameters sufficiently well as all estimates have relatively small standard errors. Second, our estimates reveal a significant difference in the magnitude of observable and unobservable shocks. The estimates of volatility of observable and unobservable shocks are 0.34 (SE=0.006) and 0.10 (SE=0.004), respectively. That is, $\frac{\sigma_U^2}{\sigma_O^2} = 0.1$. In other words, observable shocks account for about 90% of the overall variation in productivity while unobservable shocks contribute a much modest 10%. The difference in volatilities is statistically significant with a robust t-statistic of 37.

To evaluate the fit of the model, in Table 2 we report sample moments alongside moments implied by the model estimates. In the last column, we present t-statistics for the difference between the data and the model-implied moments. T-statistics are constructed using the Newey and West (1987) estimator of the variance-covariance matrix with three lags. We discuss the key moments below.

Power Law Firms size in the data is known to follow a power-law distribution (for example, Axtell (2001), Luttmer (2007), and Gabaix (2009)). In our model, the right tail of the distribution of firm size obeys a power law in the sense that for some constant η, $P(N_{j,t} \geq n) \propto n^{-\eta}$ for large values of n.[19] Here, η is the power law coefficient of the firm size distribution.

As shown in Table 2, the sample estimate of the power-law coefficient in firm capital is close to one, specifically 1.18. Consistent with the data, our model with constant returns to scale generates a fat tail distribution of firm size, and the model-implied power law virtually matches the point estimate in the data.

Firm Investment, Growth, and Size As Table 2 shows, our model is able to account well for the cross-sectional variation in firms' investment and growth observed in the data. Consistent with the data, the model generates a significant amount of dispersion in growth

[19]In our model, $K_{j,t}$ and $N_{j,t}$ are proportional to each other and are equivalent measures of firm size.

and investment rates across firms. In the model, the average investment rate declines monotonically from 23.6% for the bottom size-quintile to about 9.2% for the top size-sorted quintile. Similarly, the average growth rates vary between 12% for the small-size cohort and 0.3% for large firms matching well the corresponding sample statistics of 9.8% and 0.3% respectively.

Limited commitment allows our model to account for the robustly negative relationship between firm size and firm growth observed in the data. Equally important is the finding that the amount of unobservable shocks is modest compared to the size of observable shocks. As discussed in Section 3.4, as long as the amount of unobservable shocks is relatively small, risk sharing dominates incentive provision, which makes managerial compensation less sensitive to shocks than productivity. As a result, the manager's share in the firm, u, is negatively correlated with firm size. Because managers in small firms have a claim to a larger share of firms' cash flow, they have stronger incentives to invest than those in large firms. Hence, the negative relationship between size and u translates into a negative relationship between size and investment rate.[20]

Pay-Performance Sensitivity The level and the cross-sectional variation of pay-performance sensitivity also play an important role in identifying the amount of private information. In the data and in the model, PPS is measured in a panel regression of log growth in CEO wealth on log firm return, controlling for firm and time fixed effects. As Table 2 shows, in the data, pay-performance sensitivity varies substantially with size — small firms feature relatively low PPS while large firms are characterized by relatively high size sensitivity of managerial compensation. The empirical estimate of PPS varies from about 0.51 (SE=0.030) for the bottom quintile to 0.63 (SE=0.029) for the top size-sorted portfolio. The difference in PPS between the largest and smallest size cohorts is statistically significant

[20] Our model attributes 100% of the size-growth relationship to agency frictions. In the data, the negative size-growth relationship may be partially driven by selection and age. For tractability, our model does not provide a separate role for size and age in explaining firm growth.

with a robust t-statistics of 3.74. Our model implies similar magnitudes of PPS and a similar increasing pattern in pay-performance sensitivity across size-sorted portfolios. The model-implied spread between the top and the bottom quintile is about 0.12, which replicates the observed dispersion.[21]

The ability of the model to account for both the level and the cross-sectional variation in PPS relies crucially on the presence yet a relatively modest magnitude of unobservable shocks. At the point estimates, unobservable shocks account for about 10% of the total firm-level volatility. As explained in Section 3.4, a relatively low magnitude of unobservable shocks implies a negative relationship between firm size and manager's normalized utility, u. Recall that the sensitivity of manager's utility to productivity shocks is decreasing in u. Taken together, the two implications lead to a positive relationship between pay-performance sensitivity and firm size and allow the model to simultaneously match the level and the cross-sectional pattern in PPS observed in the data.

In our empirical analysis, we use the log-log measure of PPS because it most closely corresponds to the theoretical elasticity g_U in our model, which under the assumption of the CRRA utility is invariant to the choice of unit of denomination of utility and cash flow. As discussed above, in the log-log regression, we find a strong positive relationship between PPS and firm size. Previous literature has documented a declining pattern in level-level PPS with respect to size, that is, when PPS is measured by the regression coefficient of the dollar change in CEO wealth onto the dollar change in firm value (for example, Schaefer (1998), and Edmans, Gabaix, and Landier (2009)). Our findings do not contradict their evidence. In particular, if we follow the approach of Edmans, Gabaix, and Landier (2009) and regress the level-level measure of PPS on log of firm size, we similarly obtain a significantly negative coefficient of -0.01 (t-stat $= -13.4$).[22] The cross-sectional variation in both log-log and

[21]In the data, the positive PPS-size relationship is robust to alternative measures of firm size. For example, if we use the number of employees as a measure of firm size, we find a significant increase in PPS from 0.50 for small firms to 0.65 for large firms. Alternatively, if we sort firms on market capitalization, PPS increases monotonically from about 0.43 to 0.73 for the small and large quintiles, respectively.

[22]To replicate the Edmans, Gabaix, and Landier (2009) evidence, we use a rolling-window regression

level-level PPS is also consistent with our model. As Figure 3 shows, because the normalized utility (u) is a decreasing function of firm size, the model-implied log-log measure of PPS is increasing and the level-level measure of PPS is declining with firm size.

As Table 2 shows, overall, the model accounts well for the joint distribution of firm size, growth and managerial compensation. None of the 17 moment conditions exploited in estimation are statistically significant, and the model is not rejected by the over-identifying restrictions (the p-value of the chi-square test statistic is 0.30).

4.4 The Role of Moral Hazard

In order to better understand why the data implies a relatively small magnitude of unobservable shocks, we consider a constrained version of the model specification that assigns a larger role to moral hazard. In particular, we impose the constraint that the magnitudes of observable and unobservable shocks are equal, i.e., $\sigma_U^2 = \sigma_O^2 = 0.5\sigma^2$. Instead of evaluating the restriction directly (without re-estimating other model parameters), we give the constrained specification a fair chance to match the data and estimate the rest of its parameters by exploiting the same set of moment conditions. Table 3 presents the estimates of the constrained specification, and Table 4 reports its implications.

First notice that imposing the constraint limits the model's ability to generate a sizable cross-sectional variation in growth and investment rates. For example, our benchmark unrestricted model generates a spread of about 12% in average growth rates of firms in the bottom and the top size quintiles, matching well the 10% observed in the data. Under the constraint, the difference in average growth rates between the two firm cohorts shrinks to about 6%. These implications are quite intuitive. Keeping everything else constant, a

approach by first estimating level-level PPS for each firm in our sample by regressing the dollar change in CEO wealth on the change in market capitalization using the most recent ten-year window of data, and then run a regression of firm-level PPS on the log of firm size. The reported number is the time-series average of the estimated cross-sectional slope coefficients. If instead we run a similar regression using the log-log PPS measure, we obtain a positive estimate of 0.04 with a robust t-statistic of 10.7.

larger magnitude of unobservable shocks and a higher degree of moral hazard reduce the overall investment and growth in the economy because a significant share of capital has to be allocated to incentive provision. Therefore, investment and growth rates are substantially reduced and so is the variation in average growth rates across size-sorted portfolios. While other model parameters (eg., productivity) try to adjust and compensate the negative impact of moral hazard on growth, their adjustment is bound by the discipline imposed by other moment conditions.

Further notice the impact of moral hazard on pay-performance sensitivity. Under the constraint, the model-implied level of PPS is much larger than in the data and features no cross-sectional dispersion. As Table 4 shows, for all but the largest portfolio, the constrained specification significantly overstates size elasticity of CEO compensation. In the absence of moral hazard, CEO compensation in small (large) firms is sensitive to negative (positive) shocks but is inelastic otherwise. Introducing a relatively high degree of moral hazard makes CEO compensation of small (large) firms also respond to positive (negative) innovations and, therefore, magnifies pay-performance sensitivity. This is why the constrained specification significantly overstates the level and understates the cross-sectional variation in PPS observed in the data. Overall, allowing unobservable shocks to be relatively large leads to a significant deterioration in the model's fit as the constrained specification fails to match more than one-third of the moments exploited in estimation.

4.5 Cross-Sectional Implications

In this section, we consider an extension of our model that allows for heterogenous exposure to aggregate risks. We demonstrate three unique implications of the extended model: compared to low-exposure firms, firms with high exposure feature i) a stronger size-PPS relationship, ii) a higher investment rate, and iii) a higher level of PPS. We provide empirical evidence that supports all three implications.

Consider a cross-section of firms that differ in their exposure to aggregate shocks $dB_{Z,t}$. Extending the specification in Equation (3), we assume that the law of motion of capital stock follows

$$dK_{j,t} = K_{j,t}\left[(i_{j,t} - \delta)\, dt + \sigma_U dB_{U,t} + \sigma_O dB_{O,t} + \beta_j\left(\mu_Z dt + \sigma_Z dB_{Z,t}\right)\right], \tag{26}$$

where β_j is firm-specific exposure (beta) to aggregate risk.

Variation in aggregate risk exposure has three main implications for policy functions. First, under the optimal contract, an increase in β raises the level of the value function and expands the Pareto frontier. We illustrate this implication in Figure 8 by plotting the normalized firm value as a function of normalized utility u for three levels of aggregate risk: $\beta = 0.75$, $\beta = 1$, and $\beta = 1.25$. As the figure shows, a higher beta is associated with a higher value function and a wider range of feasible promised utility for mangers. Because, on average, the aggregate economy grows (i.e., $\mu_Z > 0$), high-β firms feature a high productivity growth ($\beta\mu_Z$), and hence, a high joint surplus of the shareholder-manager relationship. Therefore, high-β firms can support a larger range of the promised utility for managers and a higher firm value for a given level of u compared with low-β firms.

Second, the sensitivity with respect to aggregate shock, g_Z, is increasing in β. As we show in Appendix B.2, Equation (50), $g_Z = \frac{\gamma+\beta\Gamma(u)}{\gamma+\Gamma(u)}$; therefore, $\frac{\partial g_Z}{\partial \beta} = \frac{\Gamma(u)}{\gamma+\Gamma(u)}$. As long as $\Gamma(u) > 0$, $\frac{\partial g_Z}{\partial \beta} \in (0,1)$. This is the implication of optimal sharing of aggregate risks. As firm cash flow becomes more sensitive to aggregate shocks, both managers and shareholders bear part of the risk. $\frac{\partial g_Z}{\partial \beta} > 0$ because some of aggregate risk exposure is passed onto managerial compensation, and $\frac{\partial g_Z}{\partial \beta} < 1$ because shareholders retain some of the risk under the optimal contract. Note also that $\frac{\partial g_Z}{\partial \beta} = \frac{\Gamma(u)}{\gamma+\Gamma(u)}$ is increasing in $\Gamma(u)$. As the induced risk aversion of shareholder value function increases, a larger fraction of aggregate risk is passed through to managers.

Figure 9 illustrates the impact of β on the optimal policy functions g_Z (top panel), g_U

(middle panel), and investment rate, i (bottom panel). For ease of comparison across models, we re-scale u so that all three specifications have the same range of normalized utility. As the top panel shows, g_Z is increasing in β. Note also that, in general, the dispersion in g_Z is smaller than the dispersion in β, i.e., $\frac{\partial g_Z}{\partial \beta} \in (0, 1)$.[23] In the interior, when $\Gamma(u)$ is close to zero, $g_Z = \frac{\gamma + \beta\Gamma(u)}{\gamma + \Gamma(u)}$ tends towards 1, and on the boundaries, as $\Gamma(u) \to \infty$, $g_Z \to \beta$.

Third, as the bottom panel of Figure 9 shows, investment rate is increasing in β because high-β firms have high expected productivity growth, which requires a high level of investment. In order to induce a high level of investment, the elasticity of promised utility with respect to unobservable shocks, g_U, is also increasing in β, as shown in the middle panel.

We evaluate the quantitative implications of aggregate risk exposure for the equilibrium PPS-size and investment-size relationships and compare them with the data in Table 5. In the data, we sort firms into two groups — one that has high exposure to aggregate risk and the other one that features relatively low exposure. We measure aggregate risk by the market beta at the industry level. We rely on industry-level betas because variation in firm-level returns is mostly driven by idiosyncratic shocks, which hinder accurate measurement of market risk at the firm level.[24] Using 30 industry sorted portfolios, we first run a linear regression of industry excess return on the excess return of the market to estimate industry-level market risk. We then sort firms according to the market beta of the industry they belong to.[25] The high-beta cohort represents firms with the market beta above one, and the low-beta group contains firms that have less than unit exposure to the market.[26] We divide each beta cohort into three size portfolios, and in the "Data" panel of Table 5, we examine investment-size and PPS-size relationships within each beta-group. To conserve space, we only present moments of the top and bottom size terciles. The "Model" panel of Table 5 reports the corresponding model-implied moments. We set aggregate risk exposure of low-

[23]Except for a small region close to u_{MIN}, where $\frac{\partial g_Z}{\partial \beta} > 1$, because in this region, $v'(u) > 0$ and $\Gamma(u) < 0$.

[24]This is a common approach in the asset pricing literature, eg., Fama and French (1993).

[25]Industry sorted portfolios data come from the online data library maintained by Kenneth R. French.

[26]The average betas of low- and high-β industries are 0.75 and 1.22, respectively. The number of firm in each cohort is roughly the same.

and high-β firms in the model at 0.75 and 1.25, respectively, to match the average betas of low- and high-β industries in the data. All other model parameters are set at their estimated and calibrated values reported in Table 1.

First, notice that in the data and in the model, high-beta industries have higher investment rates compared with low-beta firms. This is consistent with the cross-sectional variation in investment policy functions presented in Figure 9. High aggregate risk exposure is associated with high expected productivity growth and, therefore, high rate of investment. Consequently, due to higher long-run growth, high-beta industries feature a fatter right tail in the distributions of firm size relative to low-beta industries as panel C of Table 5 shows.

Second, our model predicts that the positive PPS-size relationship should be more pronounced among high-beta firms than among low-beta firms. This is a direct implication of $\frac{\partial g_Z}{\partial \beta} = \frac{\gamma+\beta\Gamma(u)}{\gamma+\Gamma(u)} \in (0,1)$. In fact, under the optimal contract, the sensitivity of the normalized utilty with respect to aggregate shock is $\frac{d\ln u}{dB_Z} = g_Z - \beta = (1-\beta)\frac{\gamma}{\gamma+\Gamma(u)}$, which implies that positive aggregate shocks lower u if and only if $\beta > 1$. For industries with $\beta > 1$, a positive aggregate shock raises firm size and pushes firms to the left of the normalized utility space where log-log PPS is high. That is, optimal risk sharing with respect to aggregate shocks strengthens the PPS-size relationship. In contrast, for industries with $\beta < 1$, a positive aggregate shock increases u shifting firms to a region where PPS is relatively low.

This model prediction is consistent with the empirical variation of the PPS-size relationship across beta-sorted portfolios. As Panel B of Table 5 shows, in the model and in the data, the relationship between PPS and firm size for low-beta firms is virtually flat. In contrast, for high-beta firms, PPS increases significantly with size — in the data, from about 0.54 for small firms to 0.77 for large firms, and in the model, from about 0.60 to 0.79, respectively.

Further notice that in both the data and the model, high-beta firms have a higher level of PPS compared with low-beta firms, particularly among large firms. This is due to incentive provision. Recall that under the optimal contract, high-beta firms have a higher rate of investment relative to low-beta firms. Incentive compatibility requires their g_U, and hence PPS, to also be higher. Note that the estimation of our baseline model does not exploit the cross-sectional moments presented in Table 5, yet our model helps explain the implications of aggregate risk exposure for PPS-size and investment-size relationships both qualitatively and quantitatively.

4.6 Counter-Factual Exercises

In this section, we use our estimated model to conduct counter-factual experiments. First, we quantify the impact of moral hazard on CEO compensation and aggregate output by considering a setting that eliminates all moral hazard in the economy (for example, by implementing more transparent accounting rules). To carry out our analysis, we keep all preference and technology parameters at their estimated values but assume that all shocks are observable. Second, we use our model to conduct a policy experiment of imposing a limit on pay-performance sensitivity and a policy experiment of a limited CEO pay-to-firm cash flow ratio. Both policies are often proposed as potential measures to curtail inequality in managerial compensation. We demonstrate that these policies may have the unintended consequence of reducing efficiency of investment and increasing CEO pay inequality in steady state.

Impact of moral hazard on CEO compensation Moral hazard is often considered a key determinant of the level and the dynamics of CEO compensation. Our counter-factual exercise allows us to quantify the total fraction of managerial compensation that is attributed to incentive pay. To emphasize the dependence of policy functions on parameters, we denote the CEO compensation policy by $C\left(K, U|\,\Theta\right)$ and the steady-state distribution of the state variables by $\Phi\left(K, U|\,\Theta\right)$, where Θ represents the vector of parameter values of the model. We define λ_{CEOPAY} as the faction of incentive pay in total CEO compensation. Formally, we calculate the total amount of CEO compensation in the economy without moral hazard as a fraction of the total CEO pay in the economy with moral hazard and compute λ_{CEOPAY} as:

$$1-\lambda_{CEOPAY}=\frac{\int C\left(K,U|\,\Theta_0\right)\Phi\left(dK,dU|\,\Theta_0\right)}{\int C\left(K,U|\,\hat{\Theta}\right)\Phi\left(dK,dU|\,\hat{\Theta}\right)}, \tag{27}$$

where

$$\hat{\Theta}=\{\hat{\sigma}_O,\,\hat{\sigma}_U,\,\hat{\gamma}\,,\hat{\rho},\,\hat{Z}_0,\,\hat{\delta},\,\hat{\phi},\,\hat{\bar{u}}_{MIN},\,\hat{\bar{u}}_0\} \tag{28}$$

is the estimated parameter vector, and Θ_0 is obtained from $\hat{\Theta}$ by setting $\sigma_U=0$ and keeping all other parameters, including the total volatility of shocks, $\sqrt{\hat{\sigma}_U^2+\hat{\sigma}_O^2}$, the same.[27]

Our estimates imply that $\lambda_{CEOPAY}=50.4\%$. That is, moral hazard accounts for about half of the overall CEO compensation. In other words, eliminating all moral hazard allows firms to save about 50% of CEO compensation while keeping managers' utility unchanged. In our model, eliminating moral hazard makes managerial compensation contract more efficient for two reasons. First, in the presence of moral hazard, incentive provision requires CEO compensation to respond to unobservable idiosyncratic shocks. This arrangement reduces welfare because managers are risk averse. In our model with moral hazard, shocks to firm output induce variation in CEO compensation of about 18% per year, whereas perfect risk sharing implies that CEO pay grows at a constant rate over time. Note that the fact that moral hazard distorts the allocation of consumption across states of the world and limits risk sharing is true in both static and dynamic models.

Second, unique to our dynamic model, moral hazard also distorts the intertemporal allocation of managerial compensation. Under the optimal contract, the expected growth rate of continuation utility in our benchmark model with $\beta=1$ is given by (see Equation

[27] In Equation (28), we keep the time-zero productivity $\hat{Z}_0$ fixed. We use the equilibrium relationship between the marginal product of capital A and productivity Z_0 to back out $\hat{Z}_0$. See Equation (32) in Appendix A.

(56)) in Appendix C):

$$E\left[\frac{du}{u}\right]=\left[\begin{array}{c}-\frac{\rho+\kappa}{1-\gamma}\left(\left(\frac{c(u)}{u}\right)-1\right)-(i(u)-\delta+\mu_Z)+\frac{1}{2}\gamma\left(g_O^2(u)\sigma_O^2+g_U^2(u)\sigma_U^2\right)\\-\left(g_O(u)-1\right)\sigma_O^2-\left(g_U(u)-1\right)\sigma_U^2\end{array}\right] \tag{29}$$

Note that the unobservable shock B_U not only affects log-log PPS, g_U, but also distorts the intertemporal allocation of consumption by impacting the compensation and investment policies $c(u)$ and $i(u)$. Consistent with the previous literature (eg., Sannikov (2008), DeMarzo and Sannikov (2006)), moral hazard leads to a back-loaded CEO compensation package. Intuitively, delayed compensation allows shareholders to condition future compensation on realized output to provide proper incentives to invest.

To decompose the overall impact of moral hazard into its effects on risk sharing and on intertemporal allocation of consumption, respectively, let $\bar{C}\left(K,U\right)$ denote CEO compensation policy obtained from the optimal contract with moral hazard by keeping the compensation and investment policies $c(u)$ and $i(u)$ but assuming all shocks are observable so that $g_U(u)$ is chosen optimally without respecting the incentive compatibility constraint in Equation (18). Also, let $\bar{\Phi}\left(K,U\right)$ be the stationary distribution of CEO compensation under the policy $\bar{C}\left(K,U\right)$. Then, the total impact of moral hazard can be decomposed as follows:

$$\left[1-\lambda_{CEOPAY}\right]=\frac{\int\bar{C}\left(K,U\right)\bar{\Phi}\left(dK,dU\right)}{\int C\left(K,U|\,\hat{\Theta}\right)\Phi\left(dK,dU|\,\hat{\Theta}\right)}\times\frac{\int C\left(K,U|\,\Theta_0\right)\Phi\left(dK,dU|\,\Theta_0\right)}{\int\bar{C}\left(K,U\right)\bar{\Phi}\left(dK,dU\right)}. \tag{30}$$

The first term of the product on the right-hand side measures the efficiency loss due to limiting risk sharing, because moral hazard affects PPS under the contract $C\left(K,U|\,\hat{\Theta}\right)$ but not under $\bar{C}\left(K,U\right)$. The second component in Equation (30) measures the efficiency loss due to distortions in the intertemporal allocation of CEO compensation, because moral hazard affects the intertemporal allocation of consumption under the contract $\bar{C}\left(K,U\right)$ but not under $C\left(K,U|\,\Theta_0\right)$. We find that $1-\frac{\int\bar{C}(K,U)\bar{\Phi}(dK,dU)}{\int C\left(K,U|\hat{\Theta}\right)\Phi\left(dK,dU|\hat{\Theta}\right)}=21.5\%$ and $1-\frac{\int C(K,U|\Theta_0)\Phi(dK,dU|\Theta_0)}{\int\bar{C}(K,U)\bar{\Phi}(dK,dU)}=$

36.8%. Thus, our estimates imply that 50.4% of CEO compensation is the result of incentive provision. More than half of the incentive pay is compensation for distortions in the intertemporal allocation of compensation package, rather than for the additional risk managers have to bear. This decomposition highlights the importance of a dynamic setting in estimating efficiency losses — static models would miss distortions in the intertemporal allocation, which we find to be a qualitatively significant part of incentive compensation.

Impact of moral hazard on aggregate output Our general equilibrium framework also allows us to quantify the impact of moral hazard on aggregate output. We compute it as a percentage increase in output that can be achieved by eliminating all moral hazard in the economy:

$$\lambda_{OUTPUT} = \frac{\int zK\left(K,U\right)^{\alpha} N\left(K,U\right)^{1-\alpha} d\Phi\left(dK, dU \middle| \Theta_0\right)}{\int zK\left(K,U\right)^{\alpha} N\left(K,U\right)^{1-\alpha} \Phi\left(dK, dU \middle| \hat{\Theta}\right)} - 1.$$

Under the estimated parameter values of the model, we find that $\lambda_{OUTPUT} = 1.1\%$. That is, eliminating all moral hazard results in a permanent increase in aggregate output of the economy by about 1%.

It is important to note that in Section 4.2, we estimate the equilibrium marginal product of capital, A, directly because given A, all equilibrium quantities and prices that are relevant for estimation are determined independently of the productivity parameter, Z_0.[28] However, in the counter-factual exercises, all other parameters are directly taken from our estimation, except that we hold Z_0 fixed and not A. This is because Z_0 is the fundamental technology parameter, while A is an endogenous equilibrium object. The parameter Z_0 is computed from the estimated A using the equilibrium condition $A = \alpha \left(\frac{Z_0}{\mathbf{K}_0}\right)^{1-\alpha}$, where $\mathbf{K_0}$ is the time-zero aggregate capital (see Equation (32) in Appendix A).

Quantitatively, the comparative statics in general equilibrium (i.e., holding Z_0 fixed) and

[28] The only equilibrium price that is not determined is wage rate for labor. However, none of the moments we use in estimation depend on wage.

in partial equilibrium (i.e., holding A fixed) can be drastically different. Taking as given the equilibrium level of the marginal product of capital, A, a reduction in moral hazard improves efficiency of production at a firm level. This could potentially lead to an infinitely large increase in firm value because revenue is linear in A.[29] As a result, the comparative statics exercise may not even be defined in a partial equilibrium setup. However, a lower degree of moral hazard implies that firms invest more and capital accumulates faster. Hence, in general equilibrium, the marginal product of capital A must fall as the economy accumulates more capital stock. Our general equilibrium setup thus allows us to quantify the effect of moral hazard by taking into account the equilibrium relationship between the marginal product of capital and the total amount of capital in the economy.

Policy experiments In this section, we use our model as a laboratory to evaluate two policy proposals that are often suggested as means to curb inequality in CEO compensation. The first policy is to impose a limit on the CEO pay-to-worker compensation ratio. The second policy is to impose a limit on the log-log PPS of compensation contracts.

To set a benchmark for our analysis, in Figure 10, we plot the Lorenz curve of CEO pay in the 2018 data and that implied by our baseline model. Although our estimation does not target the observed inequality in CEO compensation, our model matches the data-based Lorenz curve and the Gini coefficient well. The latter is about 0.63 and 0.64 in the data and the model, respectively.

In our first policy experiment, we impose an upper bound on the compensation-to-capital ratio, $\frac{C}{K}$. Because in our model, labor and capital are proportional to each other, this specification is equivalent to imposing an upper bound on the ratio of CEO pay-to-worker compensation. Many policy proposals use the CEO pay-to-worker compensation ratio as

[29]The easiest way to see this is from the closed-form solution for the first-best firm value: $v(u) = \frac{A-\hat{i}-\phi\hat{i}^2}{\hat{r}-\hat{i}} - bu$, where $\hat{i}$ is the optimal level of investment (see Appendix B.3). Keeping A fixed, a small increase in $\hat{i}$ in the denominator can easily send the firm value to infinity.

a measure of inequality.[30] Our model provides a quantitative framework for studying the impact of those policy proposals.

Technically, we solve the constrained optimal contracting problem that imposes an upper bound on the normalized promised utility: $u \leq \tilde{u}_{MAX}$. We set $\tilde{u}_{MAX}$ to be 60% of the u_{MAX} under the optimal contract in our estimated model. We normalize CEO pay and total output in our benchmark model to 100 and report the quantitative results of our experiment in Table 6. First, note that in the economy with the restriction on the CEO-worker pay ratio, the total level of CEO compensation is about 20% lower than in our benchmark model. This is not surprising as the policy targets directly the maximum relative level of CEO pay.

Second, and surprising at first sight, the Gini coefficient in the constrained specification is significantly higher than in the benchmark model. That is, imposing an upper bound on the CEO-worker pay ratio raises inequality in CEO compensation. The increase in inequality can also be seen in Figure 11 that shows the Lorenz curve of CEO pay distribution implied by our benchmark model and in policy experiments. To understand the impact of the cap on the CEO-worker pay ratio, in Figure 12, we plot three policy functions: the log-log PPS $g_U(u)$ in the top panel, the normalized managerial compensation $c(u)$ in the middle panel, and the investment rate $i(u)$ in the bottom panel for our benchmark model and for the constrained specification. As the figure shows, the overall level of $c(u)$ in the constrained model is lower relative to the baseline. However, the log-log PPS is higher, which results in higher inequality in steady state. To understand the increase in PPS, note that according to the incentive compatibility constraint in Equation (18), the marginal cost of incentive provision is proportional to the manager's marginal utility — higher marginal utility raises the manager's incentive to use firm cash flow for private consumption. To induce an appropriate level of investment, the optimal contract responds to the cap on the CEO-worker pay ratio by raising PPS for all firms.

[30]For example, in March 2021, senators Bernie Sanders, Elizabeth Warren, Chris Van Hollen, and Edward Markey introduced the Tax Excessive CEO Pay Act in the US Senate that would increase the corporate tax for corporations whose CEO pay to median worker compensation exceeds 50 to one.

Finally, because the allocation in the benchmark model is constrained efficient, the total level of output in the model with a cap on the CEO-worker pay ratio is lower — a cap distorts investment and lowers output.

In our second experiment, we impose an upper bound on log-log PPS, g_U. Technically, we solve the constrained optimal contracting problem with $g_U \leq \tilde{g}_U$. We choose $\tilde{g}_U = 40\% \times \max_u g_U(u)$, that is, it is 40% of the highest level of $g_U(u)$ under the optimal contract. As shown in Table 6, this policy lowers the total output by 0.7%, raises the total level of CEO compensation by 22.7%, and increases CEO pay inequality in steady state.

To understand the above results, in Figure 13, we compare the policy functions of the model with an upper bound on PPS with those in our benchmark model. By design, the log-log PPS under the considered policy is lower relative to the baseline (top panel of Figure 13). To lower the marginal cost of incentive provision and induce a sufficient level of investment, the optimal contract raises managerial consumption — as shown in the middle panel, the level of $c(u)$ in the model with the restriction on PPS is higher, resulting in a higher level of overall CEO compensation, and higher inequality.

The above policy experiments highlight the importance of taking into account the endogenous response of the optimal compensation contract to policy proposals. In both cases, policies intended to curb CEO pay inequality by imposing restrictions on managerial compensation end up raising inequality in steady state due to the need of incentive provision in the presence of moral hazard.

5 Conclusion

We quantify the impact of moral hazard using a structural estimation of a dynamic general equilibrium model with agency frictions. The degree of moral hazard is defined by the relative magnitude of unobservable versus observable productivity shocks. We show that moral hazard has important implications for the cross-sectional relationships between firm size and investment, and firm size and pay-performance sensitivity. We exploit the predictions of our model and identify the amount of observable and unobservable shocks by exploiting moment conditions of the joint empirical distribution of firm size, growth and PPS. We find that the magnitude of unobservable shocks is relatively modest and accounts for about 10% of the total variation. Our estimates imply that moral-hazard induced incentives are quantitatively significant and explain 50% of managerial compensation. Our welfare analysis suggests that eliminating moral hazard would result in about 1% increase in aggregate output.

Similar to Kehoe and Levine (1993), and Albuquerque and Hopenhayn (2004), even though agents have an option to default, separation between firms and managers does not happen in equilibrium because it is associated with losses in capital and no efficiency gain. Therefore, the limited commitment specification in our model imposes feasibility constraints on the optimal contract but does not result in managers being fired in equilibrium and do not allow us to model job switching or career concerns. The absence of firing implies that variations in continuation utility must be associated with changes in future compensation. In reality, firing provides an additional way for firms to provide incentives without conditioning managerial compensation on performance within the employment relationship, which will affect the actual pay-performance sensitivity. A richer setup would allow the model to address these issues. For example, a non-trivial heterogeneity in matching quality as modeled in assortative matching models would allow job to job transition for managers and would be an interesting direction to extend the paper.

Appendix

A Equilibrium prices

In this section, we use market clearing conditions to provide expressions for two equilibrium prices: the equilibrium marginal product of capital, and the equilibrium stochastic discount factor. This will simplify the optimal contracting problem, which takes these prices as given.

The marginal product of capital Here, we solve for the equilibrium marginal product of capital. This calculation applies to both the first-best economy and economies with agency frictions.

The optimality condition for the choice of labor implies

$$(1-\alpha)\,Z_t^{1-\alpha}\left(\frac{K_{j,t}}{N_{j,t}}\right)^{\alpha}=W_t. \tag{31}$$

That is, the capital-to-labor ratio is constant across all firms. Let $\mathbf{K}_t$ denote the aggregate capital stock of the economy, the total capital stock of the economy can be computed as $\int\frac{K_{j,t}}{N_{j,t}}N_{j,t}dj=\mathbf{K}_t$. Given that $\frac{K_{j,t}}{N_{j,t}}$ is common across all firms and that the total labor supply $\int N_{j,t}dj$ is normalized to 1, the above market clearing condition for capital implies that $\frac{K_{j,t}}{N_{j,t}}=\mathbf{K}_t$ for all j. As a result, the equilibrium marginal product of capital can be computed as $\frac{d}{dK_{j,t}}Y_{j,t}=\alpha Z_t^{1-\alpha}\left(\frac{K_{j,t}}{N_{j,t}}\right)^{\alpha-1}=\alpha Z_t^{1-\alpha}\mathbf{K}_t^{\alpha-1}$.

In Section C of this appendix, we show that $\mathbf{K}_t=\frac{Z_t}{Z_0}\mathbf{K}_0$. As a result, the equilibrium marginal product of capital can be written as:

$$A=\alpha\left(\frac{Z_0}{K_0}\right)^{1-\alpha}. \tag{32}$$

In addition, Equation (31) implies that $W_t=(1-\alpha)\,Z_t^{1-\alpha}\mathbf{K}_t^{\alpha}$. After maximizing out labor, the profit function defined in Equation (2) can be written as $\pi_t A K_{j,t}$, where A is defined in Equation (32).

Because capital share is constant across all firms, firm-level output can be computed as: $Y_{j,t} = \frac{1}{\alpha} A K_{j,t}$. Integrating across all firms, aggregate output is given by:

$$\mathbf{Y}_t = \frac{1}{\alpha} A \mathbf{K_t} = Z_0^{1-\alpha} \mathbf{K}_0^{\alpha} \left(\frac{Z_t}{Z_0} \right), \tag{33}$$

where the second equality uses Equation (32). It follows that along the balance growth path, aggregate output $\mathbf{Y}_t$ grows at the same rate as Z_t.

The stochastic discount factor We conjecture here and verify below in Section B, that in the case of $\beta = 1$, the equilibrium consumption for firm owners, $C_{F,t}$, satisfies the following equation:

$$\frac{dC_{F,t}}{C_{F,t}} = \mu_Z dt + \sigma_Z dB_{Z,t}. \tag{34}$$

That is, firm owners' consumption grows at the same rate as productivity Z_t. Under the above conjecture and under the CRRA preference, an application of the Itô's lemma implies that the stochastic discount factor $\Lambda_t = e^{-\rho_F t} \left(\frac{C_{F,t}}{C_{F,0}} \right)^{-\gamma}$ follows

$$\frac{d\Lambda_t}{\Lambda_t} = -r dt - \gamma \sigma_Z dB_{Z,t}, \tag{35}$$

where the risk-free rate r is given by:

$$r = \rho_F + \gamma \mu_Z - \frac{1}{2} \gamma (1 + \gamma) \sigma_Z^2. \tag{36}$$

B Optimal Contracting

In this section, we solve the optimal contracting problem in several steps. We first establish the equivalence between the global incentive compatibility constraint (10) and the local incentive compatibility constraint (18). We then derive the HJB equation that characterizes the value function $v(u)$ for arbitrary values of β, the cash-flow exposure to aggregate shocks. Finally, we provide a closed-form solution for the first-best case and derive the HJB equation for $\beta = 1$ as special cases of the general HJB equation.

B.1 Incentive Compatibility

We summarize our main result in the following lemma.

Lemma 1. [Incentive compatibility] *A contract constructed form the allocation rule, $C(K,U)$, $I(K,U)$, $D(K,U)$, $N(K,U)$, $g_U(K,U)$, and $g_O(K,U)$ satisfies the obedience constraint* (10) *if and only if for all $t \in [0,\infty)$*

$$\frac{g_U(K_t,U_t)}{U_t^{\gamma-1}K_t} = (\rho+\kappa)\, C(K_t,U_t)^{-\gamma} H_I\left(I(K_t,U_t),K_t\right) \tag{37}$$

or in normalized terms

$$g_U(u_t) = (\rho+\kappa)\, c_t^{-\gamma} u_t^{\gamma-1} h'(i_t)\,. \tag{38}$$

Proof. We show that the obedience constraint (10) is satisfied if and only if for all $t \in [0,\infty)$

$$\left(C(K_t,U_t), I(K_t,U_t)\right) \in \arg \max_{\substack{C,I \text{ s.t} \\ C+H(I,K_t)=AK_t-D(K_t,U_t)}} \frac{\rho+\kappa}{1-\gamma} C^{1-\gamma} + \frac{1}{U_t^{\gamma-1}K_t} g_U(K_t,U_t)\, I \tag{39}$$

which along with concavity of $H(I,K)$ imply Equations (37) and (38).

To simply notation, we focus on the case when $\sigma_O = 0$, that is, all shocks are unobservable. We also omit the arguments K and U in the policy functions in the statement of the lemma and define

$$\bar{U}_t = \frac{1}{1-\gamma} U_t^{1-\gamma} = E_t \left[\int_t^{\infty} e^{-\rho(s-t)} \left(\rho + \kappa\right) C_s^{1-\gamma} ds \right], \tag{40}$$

which is the expected-utility representation of the agent's continuation utility. The Martingale representation theorem implies

$$d\bar{U}_t = \left(\rho + \kappa\right) \left[W_t - \frac{C_t^{1-\gamma}}{1-\gamma} \right] dt + G_U \left(K_t, U_t\right) \sigma_U dB_{U,t} \tag{41}$$

with

$$G_U \left(K_t, U_t\right) = \frac{g_U \left(K_t, U_t\right)}{U_t^{\gamma-1}} \tag{42}$$

and $g_U \left(K_t, U_t\right)$ being the sensitivity term in Equation (15). Suppose that $\left\{\tilde{C}_t, \tilde{I}_t\right\}_{t=0}^{\infty}$ is an alternative consumption and investment plan other than $\{C_t, I_t\}_{t=0}^{\infty}$, such that

$$\tilde{C}_t + H\left(\tilde{I}_t, K_t\right) = C_t + H\left(I_t, K_t\right) = AK_t - D_t$$

We define

$$\mathcal{G}_t^{\tilde{C},\tilde{I}} = \int_0^t e^{-(\rho+\kappa)s} \left(\rho + \kappa\right) \frac{1}{1-\gamma} \tilde{C}_s^{1-\gamma} ds + e^{-(\rho+\kappa)t} \bar{U}_t.$$

That is, $\mathcal{G}_t^{\tilde{C},\tilde{I}}$ is the time-t conditional expected utility of the agent's life-time utility if he follows plan $\left\{\tilde{C}_t, \tilde{I}_t\right\}_{t=0}^{\infty}$ over $[0, t]$ and switches to $\{C_t, I_t\}_{t=0}^{\infty}$ at t. Obviously, $\mathcal{G}_0^{\tilde{C},\tilde{I}} = \bar{U}_0$ and

$$e^{(\rho+\kappa)t} d\mathcal{G}_t^{\tilde{C},\tilde{I}} = \left(\rho + \kappa\right) \frac{1}{1-\gamma} \tilde{C}_t^{1-\gamma} dt - \left(\rho + \kappa\right) \bar{U}_t dt + d\bar{U}_t. \tag{43}$$

Let $\left\{B_{U,t}^{C,I}\right\}_{t=0}^{\infty}$ and $\left\{B_{U,t}^{\tilde{C},\tilde{I}}\right\}_{t=0}^{\infty}$ be the Itô's processes which are standard Brownian motions under the probability measures induced by $\{C_t, I_t\}_{t=0}^{\infty}$ and $\left\{\tilde{C}_t, \tilde{I}_t\right\}_{t=0}^{\infty}$, respectively, according to Girsanov theorem. Then Equation (14) implies

$$\sigma_U dB_{U,t}^{C,I} = \sigma_U dB_{U,t}^{\tilde{C},\tilde{I}} + \frac{1}{K_t} \left(\tilde{I}_t - I_t\right) dt. \tag{44}$$

Using Equations (43) and (41),

$$\begin{aligned} e^{(\rho+\kappa)t} d\mathcal{G}_t^{\tilde{C},\tilde{I}} &= \frac{\rho+\kappa}{1-\gamma}\left(\tilde{C}_t^{1-\gamma} - C_t^{1-\gamma}\right) dt + G_U\left(K_t, U_t\right) \sigma_U dB_{U,t}^{C,I} \\ &= \left[\frac{\rho+\kappa}{1-\gamma}\left(\tilde{C}_t^{1-\gamma} - C_t^{1-\gamma}\right) + G_U\left(K_t, U_t\right) \frac{1}{K_t}\left(\tilde{I}_t - I_t\right)\right] dt \\ &\quad + G_U\left(K_t, U_t\right) dB_{U,t}^{\tilde{C},\tilde{I}}. \end{aligned} \tag{45}$$

The last equality above is due to Equation (44). Suppose that Equation (39) is not satisfied and, according to Equation (42),

$$\frac{\rho+\kappa}{1-\gamma} C_t^{1-\gamma} + \frac{1}{K_t} G_U\left(K_t, U_t\right) I_t < \frac{\rho+\kappa}{1-\gamma} \tilde{C}_t^{1-\gamma} + \frac{1}{K_t} G_U\left(K_t, U_t\right) \tilde{I}_t$$

over a time interval with a positive measure. Then there exists a $\bar{t}$ such that $\left\{\mathcal{G}_t^{\tilde{C},\tilde{I}}\right\}$ is a sub-martingale over $[0, \bar{t}]$ and

$$\bar{U}_0 = \mathcal{G}_0^{\tilde{C},\tilde{I}} < E_0\left[\mathcal{G}_{\bar{t}}^{\tilde{C},\tilde{I}}\right]$$

so that $\{C_t, I_t\}_{t=0}^{\infty}$ is dominated by following $\left\{\tilde{C}_t, \tilde{I}_t\right\}_{t=0}^{\infty}$ from the beginning and switching to $\{C_t, I_t\}_{t=0}^{\infty}$ at $\bar{t}$. Conversely, if Equation (39) is satisfied for all t, there is no such profitable deviation for all $t \in [0, \infty)$. Proof of Equation (37) is straightforward. □

B.2 The HJB Equation with Arbitrary β

Taking the form of the stochastic discount factor in Equation (35) as given, we consider a firm with an arbitrary $\beta > 0$ and derive the HJB differential equation characterizing the optimal contract. The laws of motion of K, Equation (14), is written as

$$\frac{dK}{K} = \left[i\left(K, U\right) - \delta\right] dt + \sigma_O dB_O + \sigma_U dB_U + \beta\mu_Z dt + \beta\sigma_Z dB_Z. \tag{46}$$

The value function $V(K,U)$ is defined as

$$V\left(K_t,U_t\right)=\max_{\{c_t\},\{i_t\},\{g_{U,t}\},\{g_{O,t}\},\{g_{Z,t}\}} E_t\left[\int_0^\infty e^{-\kappa(t+s)}\frac{\Lambda_{t+s}}{\Lambda_t}\left(A-h\left(i_{t+s}\right)-c_{t+s}\right)K_{t+s}ds\right]$$

For any $\Delta>0$, $V(K,U)$ satisfies

$$V(K_t,U_t)=\max E_t\left[\int_0^\Delta e^{-\kappa(t+s)}\frac{\Lambda_{t+s}}{\Lambda_t}\left(A-h\left(i_{t+s}\right)-c_{t+s}\right)K_{t+s}ds+\frac{\Lambda_{t+\Delta}}{\Lambda_t}V(K_{t+\Delta},U_{t+\Delta})\right]$$

Taking the limit as $\Delta\to 0$, $V(K_t,U_t)$ must satisfy

$$e^{-\kappa t}\Lambda_t\left[A-h\left(i_t-c_t\right)\right]K_t+\mathcal{L}\left[e^{-\kappa t}\Lambda_t V(K_t,U_t)\right]=0,$$

where the operator $\mathcal{L}$ is defined as $\mathcal{L}X_t=\lim_{\Delta\to 0}\frac{1}{\Delta}E_t\left[X_{t+\Delta}-X_t\right]$, for any X_t.

Therefore, Equation (46) and the law of motion of U, Equation (15), imply that when neither limited commitment constraint is binding, $V(K,U)$ satisfies the following HJB differential equation:

$$\left(r+\kappa\right)V\left(K,U\right)=\max_{c,i,g_U,g_O,g_Z}\left\{\begin{array}{c} AK-h\left(i\right)K-cK+V_K\left(K,U\right)K\left[i-\delta+\beta\mu_Z-\gamma\beta\sigma_Z^2\right] \\ +\frac{1}{2}V_{KK}\left(K,U\right)K^2\left(\sigma_U^2+\sigma_O^2+\beta^2\sigma_Z^2\right) \\ +V_U\left(K,U\right)U\left[\begin{array}{c}\frac{\rho+\kappa}{1-\gamma}\left(1-\left(\frac{C}{U}\right)^{1-\gamma}\right) \\ +\frac{1}{2}\gamma\left(g_U^2\sigma_U^2+g_O^2\sigma_O^2+g_Z^2\sigma_Z^2\right)-\gamma g_Z\sigma_Z^2\end{array}\right] \\ +\frac{1}{2}V_{UU}\left(K,U\right)U^2\left(g_U^2\sigma_U^2+g_O^2\sigma_O^2+g_Z^2\sigma_Z^2\right) \\ +V_{K,U}\left(K,U\right)KU\left(g_U\sigma_U^2+g_O\sigma_O^2+g_Z\beta\sigma_Z^2\right)\end{array}\right\} \tag{47}$$

with the incentive constraint (38) being imposed on the maximization problem on the right hand side of Equation (47). Furthermore, according to the normalization, $v\left(\frac{U}{K}\right)=\frac{V(K,U)}{K}$, we have $V_K(K,U)=v(u)-uv'(u)$, $V_U(K,U)=v'(u)$, $V_{KK}(K,U)=\frac{1}{K}u^2v''(u)$, $V_{UU}(K,U)=\frac{1}{K}v''(u)$, and $V_{KU}(K,U)=-\frac{1}{K}uv''(u)$. Using these homogeneity conditions, we can simplify

Equation (47) as the following normalized HJB differential equation

$$0=\max_{c,i,g_O,g_U}\left\{\begin{array}{c} A-c-h(i)+v(u)\left(i-r-\kappa-\delta+\beta\mu_Z-\beta\gamma\sigma_Z^2\right) \\ +uv'(u)\left[\begin{array}{c}\frac{\rho+\kappa}{1-\gamma}\left(1-\left(\frac{c}{u}\right)^{1-\gamma}\right)-(i-\delta) \\ +\frac{1}{2}\gamma\left(g_U^2\sigma_U^2+g_O^2\sigma_O^2+g_Z^2\sigma_Z^2\right) \\ -\gamma\sigma_Z^2\left(g_Z-\beta\right)-\beta\mu_Z\end{array}\right] \\ +\frac{1}{2}u^2v''(u)\left[\left(g_U-1\right)^2\sigma_U^2+\left(g_O-1\right)^2\sigma_O^2+\left(g_Z-\beta\right)^2\sigma_Z^2\right]\end{array}\right\}. \tag{48}$$

The first-order conditions for g_O and g_Z imply

$$g_O(u)=\frac{uv''(u)}{\gamma v'(u)+uv''(u)}, \tag{49}$$

and

$$g_Z(u)=\frac{\gamma v'(u)+\beta uv''(u)}{\gamma v'(u)+uv''(u)}. \tag{50}$$

Define $\Gamma(u)=\frac{uv''(u)}{v'(u)}$ as the Arrow-Pratt measure of relative risk aversion of the value function, then we can write the above policy functions as $g_O(u)=\frac{\Gamma(u)}{\gamma+\Gamma(u)}$, and $g_Z(u)=\frac{\gamma+\beta\Gamma(u)}{\gamma+\Gamma(u)}$.

Finally, we note that under the optimal policies, the normalized continuation utility, u, follows

$$\begin{aligned}\frac{du}{u} &= \left[\begin{array}{c}-\frac{\rho+\kappa}{1-\gamma}\left(\left(\frac{c(u)}{u}\right)-1\right)-\left(i(u)-\delta+\beta\mu_Z\right) \\ +\frac{1}{2}\gamma\left(g_O^2\sigma_O^2+g_U^2\sigma_U^2+g_Z^2\sigma_Z^2\right) \\ -\left(g_O(u)-1\right)\sigma_O^2-\left(g_U(u)-1\right)\sigma_U^2-\beta\left(g_Z-\beta\right)\sigma_Z^2\end{array}\right]dt \\ &\quad +\left(g_O(u)-1\right)\sigma_O dB_O+\left(g_U(u)-1\right)\sigma_U dB_U+\left(g_Z(u)-\beta\right)\sigma_Z dB_Z.\end{aligned} \tag{51}$$

In particular, if $\beta=1$, then $g_Z-\beta=0$, and the law of motion of u does not depend on the aggregate shock $B_{Z,t}$.

B.3 The First-Best Case

We first provide the solution for the value function for the first-best case, where there is no moral hazard or limited commitment. We also assume that $\beta = 1$. The value function should still satisfy the ODE in Equation (48) except that $\sigma_U = 0$ and the incentive compatibility constraints are no longer necessary. We guess that the value function is linear: $v(u) = \bar{v} - bu$, for some constants $\bar{v}$ and b. Then, the ODE in Equation (48) can be used to determine the constants $\bar{u}$ and b.

Given the guess of the value function, it is straightforward to verify that

$$b = \frac{1}{r + \kappa + \gamma\sigma_Z^2}, \tag{52}$$

and the constant $\bar{v}$ and the optimal investment are jointly determined by the following simplified HJB equation

$$0 = \max_i A - i - \phi i^2 + \bar{v}(i - \hat{r}) \tag{53}$$

with $\hat{r} = r + \kappa + \delta - \mu_Z + \gamma\sigma_Z^2$. Let $\hat{i}$ be the first-best investment-to-capital ratio. Optimality with respect to investment implies

$$\bar{v} = 1 + 2\phi\hat{i}.$$

Using Equation (53), we know that $\hat{i}$ must satisfy $\phi\hat{i}^2 - 2\phi\hat{r}\hat{i} + (A - \hat{r}) = 0$. Therefore

$$\hat{i} = \hat{r} - \sqrt{\hat{r}^2 - \frac{1}{\phi}(A - \hat{r})}$$

For the solution to be well defined, we must have

$$\hat{r} \geq \frac{\sqrt{1 + 4A\phi} - 1}{2\phi}.$$

Intuitively, the interest rate needs to be high enough to guarantee that the value of the firm is finite.

B.4 Proof of Proposition 1

Proposition 1 is a special case of the general setup discussed above with $\beta = 1$. Setting $\beta = 1$, Equation (50) implies that $g_Z = 1$ and $g_Z - \beta = 0$. This proves the HJB Equation (21). In addition, the boundary conditions can be shown by following the argument in the proof of Lemma 1 in Ai and Li (2015).

C Aggregation

In this section, we close the model by demonstrating that in the case of $\beta = 1$, the aggregate capital stock $\mathbf{K}_t$ grows at the same rate as Z_t, and the allocations constructed from the optimal contracting problem satisfy the aggregate market clearing condition (13). We also derive an ODE that characterizes the cross-sectional distribution of promised utility that is needed for aggregation.

We define $k_{j,t} = \frac{K_{j,t}}{Z_t}$. From Equation (3), in equilibrium, $k_{j,t}$ must satisfy

$$dk_{j,t} = k_{j,t} \left[\left(i(u_{j,t}) - \delta \right) dt + \sigma^T dB_{j,t} \right], \tag{54}$$

where $i(u)$ is the optimal policy described in Proposition 1. In addition, for each j, let $u_{j,t}$ be defined as in Equation (51) where B_O and B_U are interpreted as firm-j specific Brownian motions. Given the process of $k_{j,t}$ and $u_{j,t}$, let $\Phi(u, k)$ be the invariant distribution of $(u_{j,t}, k_{j,t})$. Note that under the assumption that $\beta = 1$, the law of motion of $u_{j,t}$ does not depend on aggregate shock $B_{Z,t}$, as a result, the distribution $\Phi(u, k)$ does not depend on Z_t.[31]

We construct firm-level managerial compensation, investment, and dividend payout policies as $C_{j,t} = c(u_{j,t})k_{j,t}Z_t$, $I_{j,t} = i(u_{j,t})k_{j,t}Z_t$, and $D_{j,t} = \left[A - c(u_{j,t}) - h(i(u_{j,t}))\right] k_{j,t}Z_t$.

[31] $\Phi(u, k)$ can be characterized as the solution to a forward equation. However, an explicit characterization of this distribution is unnecessary. Numerically, it is enough to compute an appropriately defined projection of $\Phi(u, k)$ onto the space of u, which we define as the summary measure m.

We define firm-owner consumption as $C_{F,t} = (1-\alpha)\mathbf{Y}_t + \int D_{j,t}dj$. That is, firm owners consume all labor income and dividend payment. To see that this construction satisfies the resource constraint, note that

$$\begin{aligned}\int C_{j,t}dj + \int h(i_{j,t})K_{j,t}dj + C_{F,t} &= \int C_{j,t}dj + \int h(i_{j,t})K_{j,t}dj + (1-\alpha)\mathbf{Y}_t + \int D_{j,t}dj \\ &= \int AK_{j,t}dj + (1-\alpha)\mathbf{Y}_t \\ &= \mathbf{Y}_t,\end{aligned}$$

where the last line uses condition (33).

To construct the balanced growth path, we first define a summary measure that can be used to compute the market clearing conditions. For every u, define

$$m(u) = \int k\Phi(u,k)\,dk.$$

Intuitively, $m(u)$ is the total amount of k for a given firm type u. Because policy functions are homogeneous with respect to k, the summary measure contains all the information in the two dimensional distribution $\Phi(u,k)$ that is needed to compute aggregate quantities.

We can compute the aggregate managerial compensation as

$$\begin{aligned}\int C(u_{j,t}K_{j,t}, K_{j,t})\,dj &= \int c(u_{j,t})K_{j,t}dj \\ &= \int c(u_{j,t})k_{j,t}dj \times Z_t \\ &= \int kc(u)\,\Phi(u,k)\,dudk \times Z_t \\ &= \int c(u)\,\phi(u)\,du \times Z_t.\end{aligned}$$

Note that the term $\int c(u)\,\phi(u)\,du$ is time invariant. As a result, aggregate managerial compensation is proportional to Z_t. Similarly, aggregate capital stock is

$$\int K_{j,t}dj = \int k_{j,t}dj \times Z_t = \int m(u)\, du \times Z_t.$$

It follow that $\mathbf{K_t}$ grows at the same rate as Z_t, which verifies our conjecture in Appendix A that $\mathbf{K}_t = \mathbf{K}_0 \frac{Z_t}{Z_0}$. Now we can use Equation (33) to compute the aggregate output as

$$\mathbf{Y}_t = \frac{1}{\alpha} Z_0^{1-\alpha} \mathbf{K}_t = \frac{1}{\alpha} Z_0^{1-\alpha} \int m(u)\, du \times Z_t. \tag{55}$$

The above equation implies that $\mathbf{Y}_t$ also grows at the same rate as Z_t. We can similarly prove that $C_{F,t}$ grows at the same rate as Z_t, which verifies Equation (34) in Appendix A.

Finally, we present a differential equation that describes the summary measure $m(u)$. According to Equation (51), when $\beta = 1$, $g_Z - \beta = 0$, the law of motion of the normalized continuation utility satisfies Equation (22), where

$$\mu_u(u) = \begin{bmatrix} -\frac{\rho+\kappa}{1-\gamma}\left(\left(\frac{c(u)}{u}\right) - 1\right) - (i(u) - \delta + \mu_Z) + \frac{1}{2}\gamma\left(g_O^2(u)\sigma_O^2 + g_U^2(u)\sigma_U^2\right) \\ -\left(g_O(u) - 1\right)\sigma_O^2 - \left(g_U(u) - 1\right)\sigma_U^2 \end{bmatrix}. \tag{56}$$

Therefore, numerically, we can compute the summary measure $m(u)$ by solving the following forward equation:[32]

$$\begin{aligned} 0 \;=\;& \bar{m}(u) - (\kappa - (i(u) - \delta + \mu_Z))\, m(u) - \frac{d}{du}\left[m(u)\mu_u(u) + u\left(\left(g_O(u) - 1\right)\sigma_O^2 + \left(g_U(u) - 1\right)\sigma_U^2\right)\right] \\ &+ \frac{1}{2}\frac{d^2}{du^2}\left(m(u)u^2\left(\left(g_O(u) - 1\right)^2\sigma_O^2 + \left(g_U(u) - 1\right)^2\sigma_U^2\right)\right). \end{aligned}$$

[32]For a detailed derivation of the summary measure for a similar problem, see Ai, Kiku, Li, and Tong (2021).

D Details of the Counter-Factual Exercise

In the decomposition in Equation (30), when constructing $\int C(K, U|\Theta_0)\,\bar{\Phi}(dK, dU|\Theta_0)$, we let $C(K, U|\Theta_0)$ be the optimal consumption policy function for the case with no moral hazard. We compute the stationary distribution $\bar{\Phi}(dK, dU|\Theta_0)$ using the law of motion of K in Equation (46) and the law of motion of u in Equation (51) where $i(u)$ is the optimal policy function without moral hazard, but g_U and g_O are the optimal policies with moral hazard. That is, we let moral hazard distort the intertemporal allocation of consumption but not the current-period consumption.

E Construction of CEO Wealth

Executive compensation data over the 1992–2019 time period come from the Standard & Poor's ExecuComp database that contains information on salary, bonuses, stock and options grants, and other forms of payouts that are reported by companies according to the SEC reporting requirements. We define CEO wealth at the end of fiscal year t as:

$$\begin{aligned} \mathrm{W_t} &= \mathrm{SALARY_t + BONUS_t + ALLOTHTOT_t + OTHANN_t + SHRS\_UNVEST\_VAL_t} \\ &+ \mathrm{V_{stock,t} + V_{option,t} + Div_t + TradingGains_t + PV_t(future\_payouts)} \end{aligned} \tag{57}$$

where the variables is all capital letters correspond to ExecuComp data definitions, in particular $\mathrm{SALARY_t}$ and $\mathrm{BONUS_t}$ are the base salary and bonus earned by an executive in fiscal year t, $\mathrm{ALLOTHTOT_t}$ accounts for severance payments, tax reimbursements, 401K contributions, signing bonuses, etc., $\mathrm{OTHANN_t}$ includes other miscellaneous items, and $\mathrm{SHRS\_UNVEST\_VAL_t}$ is the market value of restricted shares held by an executive as of fiscal year end. $\mathrm{V_{stock,t}}$ and $\mathrm{V_{option,t}}$ are the market values of shares and stock options owned, $\mathrm{Div_t}$ and $\mathrm{TradingGains_t}$ are dividends and net revenue from stock trading, and $\mathrm{PV_t(future\_payouts)}$ is the present value of future payoffs.

The market value of stock holdings is computed by multiplying the reported number of shares owned ($\mathrm{SHROWN\_TOT_t}$) by the end-of-fiscal year price ($\mathrm{p_t}$): $\mathrm{V_{stock,t}} = \mathrm{SHROWN\_TOT_t} * \mathrm{p_t}$. Different approaches have been employed in the literature to measure of the market value of stock options owned by an executive. One straightforward approach considered in Aggarwal and Samwick (1999), and Clementi and Cooley (2009) is to rely on the reported value of the in-the-money vested and unvested options ($\mathrm{OPT\_UNEX\_EXER\_EST\_VAL} + \mathrm{OPT\_UNEX\_UNEXER\_EST\_VAL}$). This measure, however, does not account for the value of exercisable and un-exercisable options that are currently out-of-the-money. To address this limitation, we consider an alternative measure proposed by Himmelberg and Hubbard (2000) and also considered in Clementi and Cooley (2009). In particular, we assume that none of the options granted in a current year are exercisable immediately, we also assume that 25 percent of the options vest each year. Under these assumptions,

$$\begin{aligned} \mathrm{V_{option,t}} &= \mathrm{OPTION\_AWARDS\_BLK\_VALUE_t} + (1 - 0.25) * \mathrm{OPT\_UNEX\_UNEXER\_NUM_{t-1}} * \mathrm{v_{un,t}} \\ &+ (0.25 * \mathrm{OPT\_UNEX\_UNEXER\_NUM_{t-1}} + \mathrm{OPT\_UNEX\_EXER\_NUM_{t-1}} - \mathrm{OPT\_EXER\_NUM_t}) * \mathrm{v_{ex,t}} \end{aligned} \quad (58)$$

where $\mathrm{OPT\_UNEX\_UNEXER\_NUM_{t-1}}$ and $\mathrm{OPT\_UNEX\_EXER\_NUM_{t-1}}$ are the numbers of un-exercised un-exercisable and un-exercised exercisable options at the beginning of fiscal year t, $\mathrm{OPTION\_AWARDS\_BLK\_VALUE_t}$ is the Black-Scholes value of stock options granted, and $\mathrm{OPT\_EXER\_NUM_t}$ is the number of options exercised in year t. Following Himmelberg and Hubbard (2000), we assume that the strike prices of exercisable and un-exercisable options are $0.1\mathrm{p_t} + 0.3\mathrm{p_{t-1}} + 0.6\mathrm{p_{t-2}}$ and $0.6\mathrm{p_t} + 0.3\mathrm{p_{t-1}} + 0.1\mathrm{p_{t-2}}$, respectively. To compute the Black-Scholes values of exercisable and un-exercisable options ($\mathrm{v_{ex,t}}$ and $\mathrm{v_{un,t}}$), we assume that exercisable and un-exercisable options, on average, mature in three and five years, respectively. Correspondingly, we use 3- and 5-year constant maturity Treasury bond rates to proxy for the risk-free rate. We estimate conditional volatility using realized variances of monthly equity returns over the previous three years. Realized variances and dividend yields are computed using monthly CRSP data, and all computations are adjusted for stock splits.

We estimate the net revenue from stock trading following the approach of Clementi and Cooley (2009). In particular, we solve for the net amount of shares purchased in fiscal year t ($\mathrm{NetPurchase_t}$) by exploiting the law of motion of shares owned ($\mathrm{SHROWN\_TOT_t}$):

$$\mathrm{SHROWN\_TOT_t} = \mathrm{SHROWN\_TOT_{t-1}} + \mathrm{OPT\_EXER\_NUM_t} + \mathrm{Vest_t} + \mathrm{NetPurchase_t} \tag{59}$$

where $\mathrm{OPT\_EXER\_NUM_t}$ is the number of shares acquired on option exercise, and $\mathrm{Vest_t}$ is the number of vested shares. The latter is obtained using the evolution of the restricted share holdings, $\mathrm{STOCK\_UNVEST\_NUM_t}$,

$$\mathrm{STOCK\_UNVEST\_NUM_t} = \mathrm{STOCK\_UNVEST\_NUM_{t-1}} + \mathrm{RSTKGRNT_t}/\mathrm{p_t} - \mathrm{Vest_t} \tag{60}$$

where $\mathrm{RSTKGRNT_t}$ is the value of restricted stock granted in year t and $\mathrm{p_t}$ is the share price. If the net amount of shares purchased is greater that zero ($\mathrm{NetPurchase_t} > 0$), we set the net revenue from stock trading to zero; otherwise, we estimate the net revenue as:

$$\mathrm{TradingGains_t} = \max\{0, \ -\mathrm{NetPurchase_t} * \mathrm{p_{av,t}} - \mathrm{Cost_t}\} \tag{61}$$

where $\mathrm{p_{av,t}}$ is the average price during fiscal year t, and $\mathrm{Cost_t}$ is the cost of option exercise. That is, we impose that the net trading profit is non-negative. To estimate the cost of exercising options, we solve for the strike price using the information on the value realized from option exercise (OPT_EXER_VAL) and the amount of exercised options ($\mathrm{OPT\_EXER\_NUM_t}$), assuming that options are exercised when the stock price is at its maximum, i.e.,

$$\mathrm{OPT\_EXER\_VAL_t} = (\mathrm{p_{max,t}} - \mathrm{Strike_t}) * \mathrm{OPT\_EXER\_NUM_t} \tag{62}$$

$$\mathrm{Cost_t} = \mathrm{OPT\_EXER\_NUM_t} * \mathrm{Strike_t} \tag{63}$$

where $\mathrm{p_{max,t}}$ is the maximum share price in year t.

Finally, we estimate the present value of future payouts as:

$$\begin{aligned} \mathrm{future\_payout} &= \mathrm{SALARY_t} + \mathrm{BONUS_t} + \mathrm{ALLOTHTOT_t} - \mathrm{ALLOTHPD_t} \qquad (64) \\ &+ \mathrm{Div_t} + \mathrm{RSTKGRNT_t} + \mathrm{OPTION\_AWARDS\_BLK\_VALUE_t} \\ \mathrm{PV_t(future\_payouts)} &= \frac{1}{r}\mathrm{future\_payout} \qquad (65) \end{aligned}$$

where $\text{ALLOTHTOT}_{\text{t}} - \text{ALLOTHPD}_{\text{t}}$ includes 401K contributions and life insurance premiums, $\text{RSTKGRNT}_{\text{t}}$ and $\text{OPTION\_AWARDS\_BLK\_VALUE}_{\text{t}}$ are the value of restricted stock grants and the Black-Scholes value of options granted during fiscal year t, respectively, and r is the discount rate.

References

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

Ai, Hengjie, Dana Kiku, Rui Li, and Jincheng Tong, 2021, A unified model of firm dynamics with limited commitment and assortative matching, *Journal of Finance* 76, 317–356.

Ai, Hengjie, and Rui Li, 2015, Investment and CEO Compensation under Limited Commitment, *Journal of Financial Economics.* 116, 452–472.

Albuquerque, Rui, and Hugo Hopenhayn, 2004, Optimal Lending Contracts and Firm Dynamics, *Review of Economic Studies* 71, 285–315.

Atkeson, Andrew, and Robert Lucas, 1992, On Efficient Distribution with Private Information, *Review of Economic Studies* 59, 427–453.

Axtell, Robert L., 2001, Zipf Distribution of U.S. Firm Sizes, *Science* 293, 1818–1820.

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

Chi, Chang-Koo, and Kyoung Jin, 2017, The impact of firm size on dynamic incentives and investment, *Rand Journal of Economics* 48, 147–177.

Clementi, Gian Luca, and Thomas F. Cooley, 2009, Executive Compensation: Facts, *NBER Working Paper No. 15426.*

DeMarzo, Peter, Michael Fishman, Zhiguo He, and Neng Wang, 2012, Dynamic Agency and q Theory of Investment, *Journal of Finance* 67, 2295–2340.

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

Di Tella, Sebastian, and Yuliy Sannikov, 2016, Optimal Asset Management Contracts with Hidden Savings, *Working paper, Stanford Univerity.*

Edmans, Alex, and Xavier Gabaix, 2016, Executive Compensation: A Modern Primer, *Journal of Economic Literature* 54, 1232–1287.

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

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

Fama, Eugene, and Kenneth French, 1993, Common Risk Factors in the Returns on Stocks and Bonds, *Journal of Financial Economics* 33, 3–56.

Feng, Felix, 2018, Dynamic Compensation under Uncertainty Shocks and Limited Commitment, *Working paper, University of Notre Dame.*

Feng, Felix, and Mark Westerfield, 2018, Dynamic Asset Allocation with Hidden Volatility, *Working paper, University of Notre Dame.*

Fudenberg, Drew, Bengt Holmstrom, and Paul Milgrom, 1990, Short-Term Contracts and Long-Term Agency Relationships, *Journal of Economic Theory* 51, 1–31.

Gabaix, Xavier, 2009, Power Laws in Economics and Finance, *Annual Review of Economics* 1, 255–294.

Gayle, George-Levi, and Rober A. Miller, 2015, Identifying and Testing Models of Managerial Compensation, *Review of Economic Studies* 82, 1074–1118.

Gryglewicz, Sebastian, Simon Mayer, and Erwan Morellec, 2020, Agency Conflicts and Short- vs Long-Termism in Corporate Policies, *Journal of Financial Economics* 136, 718–742.

Hackbarth, Dirk, Alejandro Rivera, and Tak-Yuen Wong, 2021, Optimal Short-Termism, *Management Science* forthcoming.

Hartman-Glaser, Barney, Hanno Lustig, and Mindy Xiaolan, 2019, Capital Share Dynamics When Firms Insure Workers, *Journal of Finance* 74, 1707–1751.

Hartman-Glaser, Barney, Simon Mayer, and Konstantin Milbradt, 2018, Corporate Liquidity Management under Moral Hazard, *Working Paper, University of California, Los Angeles.*

Hartman-Glaser, Barney, Tomasz Piskorski, and Alexei Tchistyi, 2012, Optimal securitization with moral hazard, *Journal of Financial Economics* 104, 186–202.

Hayashi, Fumio, 1982, Tobin's Marginal q and Average q: A Neoclassical Interpretation, *Econometrica* 50, 213–224.

He, Zhiguo, 2009, Optimal Executive Compensation when Firm Size Follows Geometric Brownian Motion, *Review of Financial Studies* 22, 859–892.

Himmelberg, Charles P., and R. Glenn Hubbard, 2000, Incentive Pay and the Market for CEOs: An Analysis of Pay-for-Performance Sensitivity, *Working paper, University of Minnesota.*

Hoffmann, Florian, and Sebastian Pfeil, 2010, Reward for Luck in a Dynamic Agency Model, *Review of Financial Studies* 23(9), 3329–3345.

Kehoe, Timothy, and David Levine, 1993, Debt-Constrained Asset Markets, *Review of Economic Studies* 60, 865–888.

King, Robert G., and Sergio T. Rebelo, 1999, Resuscitating Real Business Cycles, in *Handbook of Macroeconomics* (Elsevier,).

Kiyotaki, Nobuhiro, and John Moore, 1997, Credit Cycles, *Journal of Political Economy* 105, 211–284.

Kydland, Finn, and Edward Prescott, 1982, Time to Build and Aggregate Fluctuations, *Econometrica* 50, 1345–1370.

Leung, Raymond, 2017, Continuous-Time Principal-Agent Problem with Drift and Stochastic Volatility Control, *Working paper, CKGSB.*

Li, Rui, 2017, Dynamic Agency with Persistent Observable Shocks, *Journal of Mathematical Economics* 71, 74–91.

Li, Shaojin, and Toni Whited, 2016, Capital Reallocation and Adverse Selection, *Working paper, University of Michigan.*

Li, Shaojin, Toni Whited, and Yufeng Wu, 2016, Collateral, Taxes, and Leverage, *Review of Financial Studies* 29, 1453–1500.

Lustig, Hanno, Chad Syverson, and Stijn Van Nieuwerburgh, 2011, Technological Change and the Growing Inequality in Managerial Compensation, *Journal of Financial Economics* 99, 601–627.

Luttmer, Erzo, 2007, Selection, Growth, and the Size Distribution of Firms, *Quarterly Journal of Economics* 122, 1103–1144.

Luttmer, Erzo, 2011, On the Mechanics of Firm Growth, *Review of Economic Studies* 78, 1042–1068.

Margiotta, M. M., and R. A. Miller, 2000, Managerial Compensation and the Cost of Moral Hazard, *International Economic Review* 41, 669–719.

McFadden, Daniel, 1989, A Method of Simulated Moments for Estimation of Discrete Response Models Without Numerical Integration, *Econometrica* 57, 995–1026.

Mirrlees, James, 1974, *Notes on Welfare Economics, Information, and Uncertainty,* American Elsevier Pub. Co, pp. 243–258 in M. Balch, D. McFadden and S. Wu (Eds), Essays in Economic Behavior under Uncertainty edn.

Newey, Whitney, and Kenneth West, 1987, A Simple, Positive Semi-Definite, Heteroscedasticity and Autocorrelation Consistent Covariance Matrix, *Econometrica* 55, 703–708.

Nikolov, Boris, and Lukas Schmid, 2016, Testing Dynamic Agency Theory via Structural Estimation, *Working paper, Duke University.*

Nikolov, Boris, Lukas Schmid, and Roberto Steri, 2019, Dynamic Corporate Liquidity, *Journal of Financial Economics* 132, 76–102.

Nikolov, Boris, and Toni Whited, 2014, Agency Conflicts and Cash: Estimates from a Structural Model, *Journal of Finance* 69, 1883–1921.

Pakes, Ariel, and David Pollard, 1989, Simulation and the Asymptotics of Optimization Estimators, *Econometrica* 57, 1027–1057.

Peters, Ryan H., and Lucian A. Taylor, 2017, Intangible Capital and the Investment-Q Belation, *Journal of Financial Economics* 123, 251–272.

Piskorski, Tomasz, and Mark Westerfield, 2016, Optimal dynamic contracts with moral hazard and costly monitoring, *Journal of Economic Theory* 166, 242–281.

Rouwenhorst, K. Geert, 1995, *Asset Pricing Implications of Equilibrium Business Cycle Models.* (Princeton University Press).

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

Schaefer, Scott, 1998, The Dependence of Pay-Performance Sensitivity on the Size of the Firm, *Review of Economics and Statistics* 80, 436–443.

Sun, Qi, and Mindy Xiaolan, 2019, Financing Intangible Capital, *Journal of Financial Economics* 133, 564–588.

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

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

Williams, Noah, 2011, Persistent Private Information, *Econometrica* 79, 1233–1275.

Xiaolan, Mindy, 2014, Who Bears Firm-Level Risk? Implications for Cash Flow Volatility, *Working paper, University of Texas Austin.*

Zhu, John, 2013, Optimal Contracts with Shirking, *Review of Economic Studies* 80, 812–839.

Table 1

Model Parameters

Panel A: Estimated Parameters

Parameter	Estimate
σ_O	0.337 (0.006)
σ_U	0.099 (0.004)
γ	0.913 (0.072)
ρ	0.004 (0.001)
A	0.168 (0.001)
δ	0.066 (0.001)
ϕ	0.970 (0.055)
$\bar{u}_{MIN}$	0.035 (0.003)
$\bar{u}_0$	0.196 (0.079)

Panel B: Calibrated Parameters

Parameter	Value
σ_Z	0.030
μ_Z	0.015
ρ_F	0.030
κ	0.050
α	0.330

Table 1 presents the estimates of the model parameters and their standard errors in parentheses (Panel A) and the values of the calibrated parameters (Panel B). The estimated parameters include volatility of observable and unobservable shocks (σ_O and σ_U, respectively), risk aversion (γ), the discount rate of the manager (ρ), the total factor productivity (A), the capital depreciation rate (δ), the adjustment cost parameter (ϕ), and parameters that determine the outside option of managers and the initial normalized utility ($\bar{u}_{MIN}$ and $\bar{u}_0$, respectively). The set of calibrated parameters consists of volatility of aggregate shocks (σ_Z), the aggregate growth rate (μ_Z), the discount rate of the shareholders (ρ_F), the death rate of managers (κ), and the capital share (α).

Table 2
Sample and Model-Implied Moments

Moments		Data	Model	t-stat(Diff)
Power Law		1.184	1.177	0.38
CS-Std of Growth Rates		0.399	0.343	1.88
Growth:	P1 (Small)	0.098	0.122	−0.63
	P2	0.051	0.071	−1.00
	P3	0.027	0.035	−0.81
	P4	0.016	0.015	0.26
	P5 (Large)	0.003	0.003	0.04
I/K:	P1 (Small)	0.186	0.236	−1.19
	P2	0.157	0.171	−1.12
	P3	0.140	0.126	1.65
	P4	0.113	0.104	0.69
	P5 (Large)	0.086	0.092	−0.44
PPS:	P1 (Small)	0.511	0.544	−0.92
	P2	0.517	0.475	1.08
	P3	0.533	0.521	0.43
	P4	0.590	0.582	0.22
	P5 (Large)	0.632	0.658	−0.37

Table 2 reports moments in the data and in the model along with the robust t-statistics for the difference between sample- and model-implied moments. The set of moments consists of the power law in firm size, the cross-sectional standard deviation of firms' growth rates, the median growth and investment rates, and executive pay-performance sensitivity for the cross section of size-sorted portfolios (P1–P5). Growth rates are measured in logs, PPS is measured in a panel regression of log growth in CEO wealth on log firm return, controlling for firm and time fixed effects, and the power law is estimated using firms in the top size-sorted portfolio. The data are annual, measured in real terms and cover the period from 1992 to 2019.

Table 3

Estimates of the Constrained Model Specification

Parameter	Estimate
σ	0.310 (0.0027)
γ	0.932 (0.0145)
ρ	0.005 (0.0002)
A	0.223 (0.0006)
δ	0.117 (0.0009)
ϕ	0.229 (0.0091)
$\bar{u}_{MIN}$	0.012 (0.0008)
$\bar{u}_0$	0.254 (0.0631)
σ_O^2	$0.5\sigma^2$
σ_U^2	$0.5\sigma^2$

Table 3 presents the estimates of the constrained model specification and their standard errors in parentheses. The estimated parameters include total volatility (σ), risk aversion (γ), the discount rate of the manager (ρ), the total factor productivity (A), the capital depreciation rate (δ), the adjustment cost parameter (ϕ), and parameters that determine the outside option of managers and the initial normalized utility ($\bar{u}_{MIN}$ and $\bar{u}_0$, respectively). The bottom rows show the constrains imposed on volatilities of observable and unobservable shocks (σ_O and σ_U, respectively). The remaining parameters are calibrated as in Panel B of Table 1.

Table 4
Model-Implied Moments of the Constrained Specification

Moments		Data	Model	t-stat(Diff)
Power Law		1.184	1.292	−2.39
CS-Std of Growth Rates		0.399	0.295	2.32
Growth:	P1 (Small)	0.098	0.067	0.84
	P2	0.051	0.056	−0.27
	P3	0.027	0.023	0.46
	P4	0.016	0.008	1.65
	P5 (Large)	0.003	0.003	0.02
I/K:	P1 (Small)	0.186	0.195	−0.24
	P2	0.157	0.180	−1.60
	P3	0.140	0.140	0.05
	P4	0.113	0.120	−0.55
	P5 (Large)	0.086	0.118	−1.85
PPS:	P1 (Small)	0.511	0.718	−2.41
	P2	0.517	0.689	−2.32
	P3	0.533	0.701	−2.43
	P4	0.590	0.713	−2.19
	P5 (Large)	0.632	0.722	−1.14

Table 4 shows the implications of the constrained model specification detailed in Table 3. We report moments in the data and in the model, and robust t-statistics for the difference between sample- and model-implied moments. The set of moments consists of the power law in firm size, the cross-sectional standard deviation of firms' growth rates, the median growth and investment rates, and executive pay-performance sensitivity for the cross section of size-sorted portfolios (P1–P5). Growth rates are measured in logs, PPS is measured in a panel regression of log growth in CEO wealth on log firm return, controlling for firm and time fixed effects, and the power law is estimated using firms in the top size-sorted portfolio. The data are annual, measured in real terms and cover the period from 1992 to 2019.

Table 5

Cross-Sectional Implications

Panel A: Investment Rates

	Data			Model		
Firm Size	Low-β	High-β	High−Low	Low-β	High-β	High−Low
Small	0.116	0.150	0.034 (2.65)	0.196	0.227	0.031
Large	0.069	0.127	0.057 (2.63)	0.077	0.110	0.033
Large−Small	−0.047	−0.023		−0.120	−0.118	
	(−2.74)	(−1.71)				

Panel B: Pay-Performance Sensitivity

	Data			Model		
Firm Size	Low-β	High-β	High−Low	Low-β	High-β	High−Low
Small	0.530	0.535	0.004 (0.24)	0.532	0.603	0.071
Large	0.550	0.772	0.222 (4.78)	0.545	0.794	0.249
Large−Small	0.020	0.237		0.013	0.191	
	(0.57)	(4.98)				

Panel C: Power Law

Data			Model		
Low-β	High-β	High−Low	Low-β	High-β	High−Low
1.64	1.25	−0.39	1.30	0.99	−0.31
		(−2.80)			

Table 5 presents variation in investment growth rates (Panel A), pay-performance sensitivity (Panel B), and power-law exponent (Panel C) across 2x3 portfolios sorted by exposure to aggregate risks (β) and size. To conserve space, the statistics of the mid-size portfolios are omitted. In the data, firm exposure to aggregate risk is measured by the industry-level market beta estimated using monthly excess returns over the 1992–2019 period. Robust t-statistics for the cross-sectional differences are reported in parentheses. Growth rates are measured in logs, PPS is measured in a panel regression of log growth in CEO wealth on log firm return, controlling for firm and time fixed effects, and the power law is estimated using firms in the top size-sorted portfolio. The sample moments are based on annual data that are measured in real terms and cover the period from 1992 to 2019.

Table 6

Policy Experiments

	Benchmark	Bound on CEO pay ratio	Bound on PPS
Output	100	99.85	99.29
CEO pay	100	79.64	122.67
Gini coefficient	0.6385	0.7477	0.6893

Table 6 presents the level of output, CEO pay, and the Gini coefficient of the distribution of CEO compensation in our benchmark model and in policy experiments. We normalize the level of output and CEO pay in the baseline model to 100.

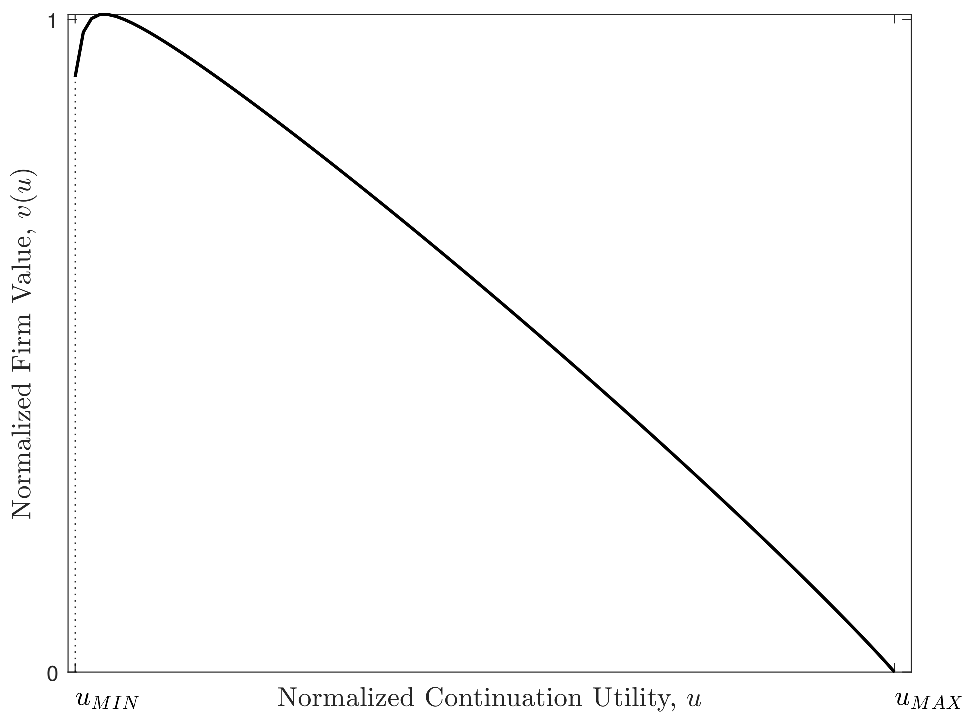


Figure 1. Normalized Value Function

Figure 1 plots the normalized value function. The horizontal axis represents the normalized continuation utility of the manager, u, and the vertical axis represents the normalized firm value. u_{MIN} and u_{MAX} are the lower and upper bounds of u respectively.

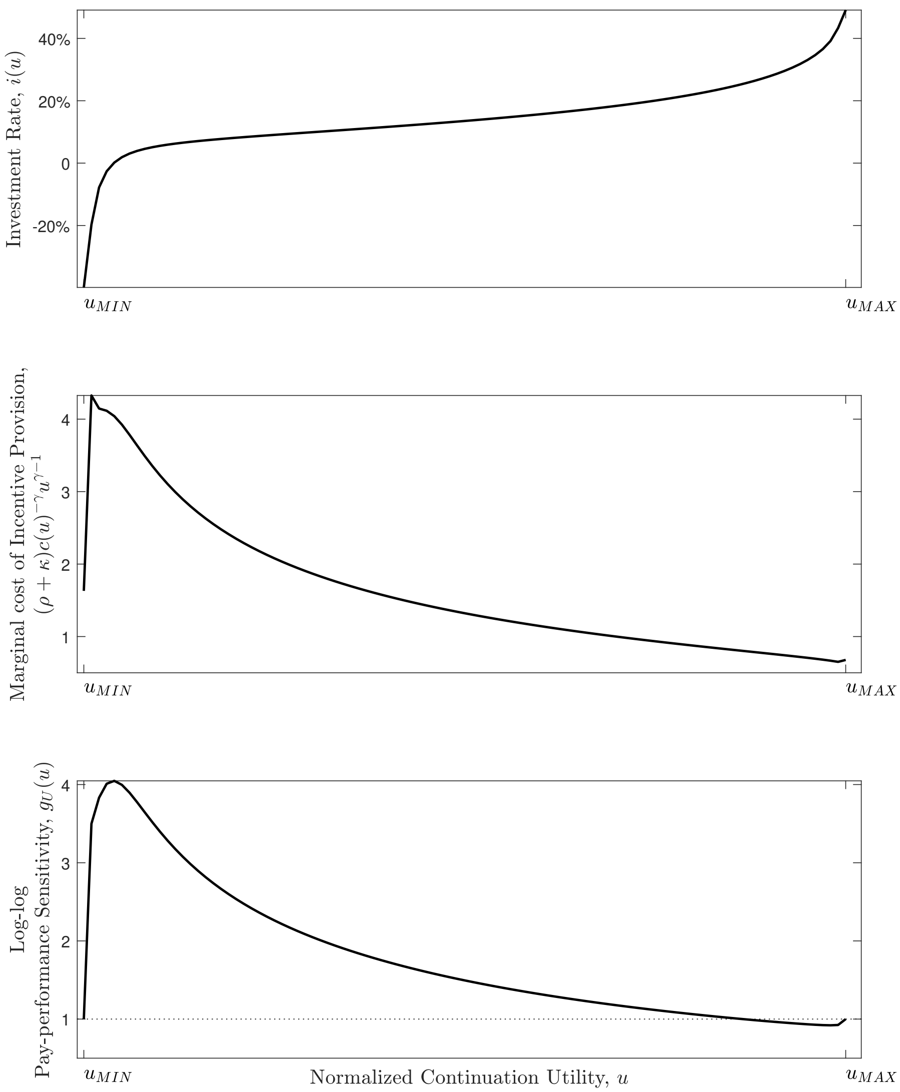


Figure 2. Investment Rate and Marginal Cost of Incentive Provision

Figure 2 plots the investment-to-capital ratio $i(u)$ (top panel), the marginal cost of incentive provision, $(\rho+\kappa)c(u)^{-\gamma}u^{\gamma-1}$ (middle panel), and the log-log PPS with respect to unobservable shocks, $g_U(u)$ (bottom panel), under the optimal contract as functions of the normalized utility.

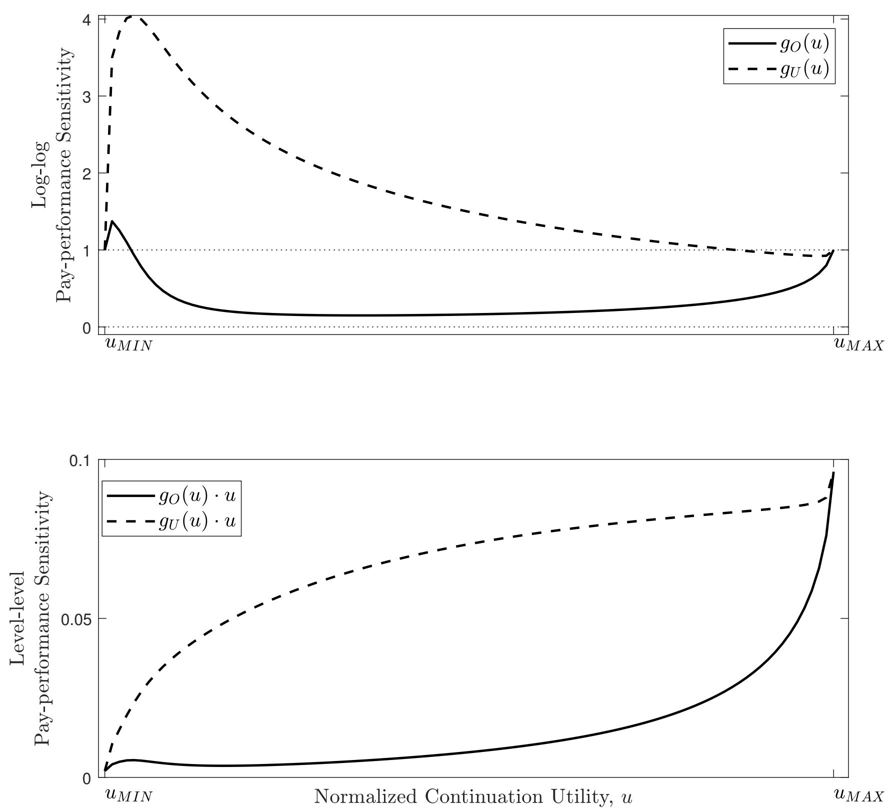


Figure 3. Elasticity of Continuation Utility

The top panel of Figure 3 plots the policy functions for the log-log PPS with respect to observable shocks, $g_O(u)$, and unobservable shocks, $g_U(u)$ (solid and dashed lines, respectively). The bottom panel presents the corresponding functions for the level-level PPS, $g_O(u) \cdot u$ and $g_U(u) \cdot u$.

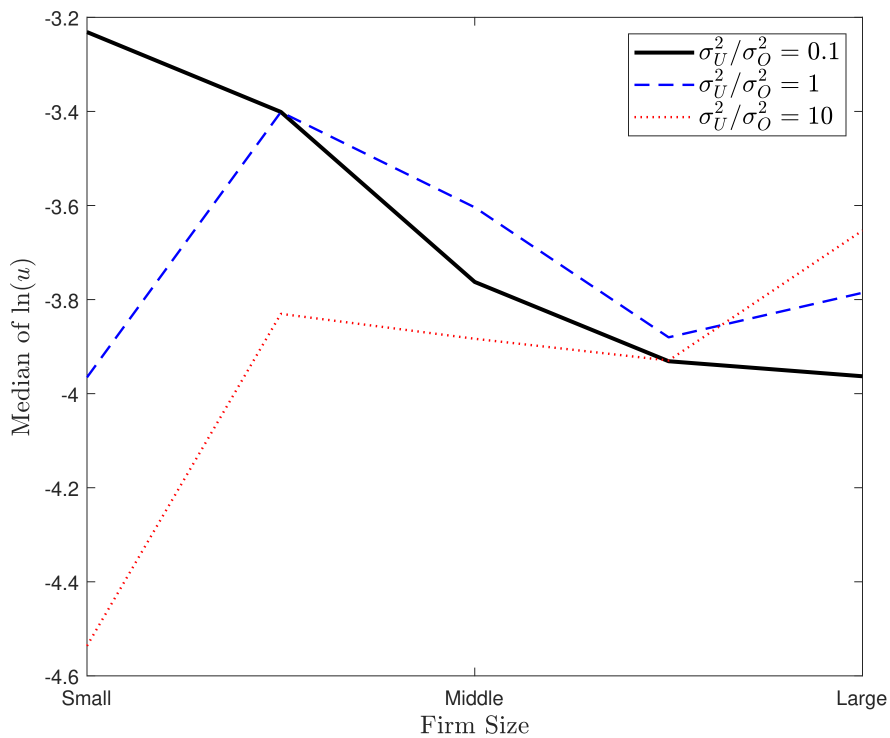


Figure 4. Manager's Equity Share and Firm Size

Figure 4 plots the equilibrium relationship between the manager's equty share (in logs) and firm size under different assumptions about the relative magnitude of unobservable and observable shocks, $\frac{\sigma_U^2}{\sigma_O^2}$.

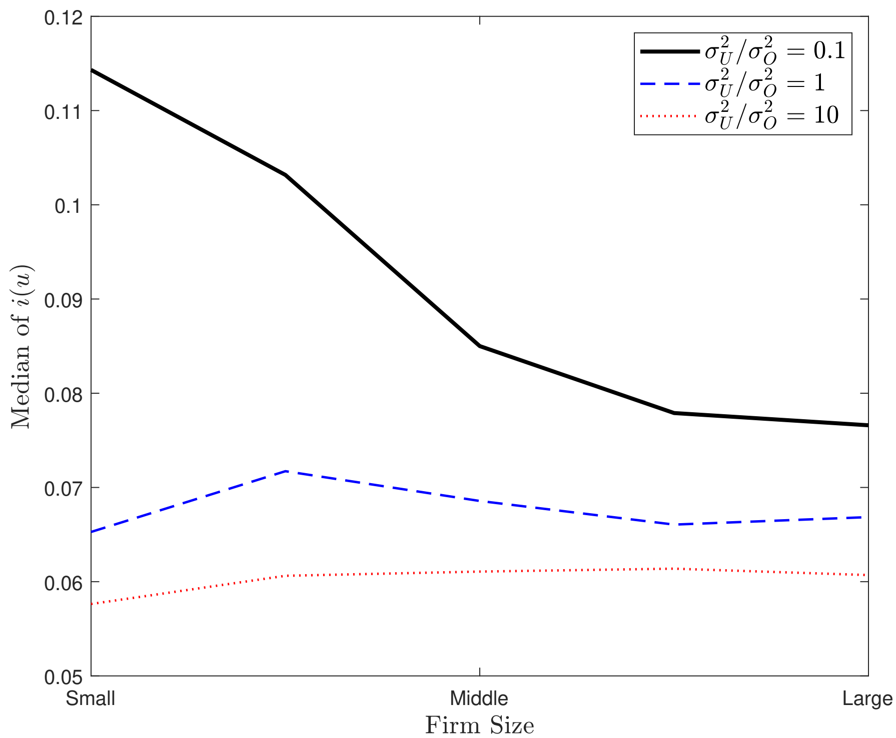


Figure 5. Investment and Firm Size

Figure 5 plots the equilibrium relationship between the average investment-to-capital ratio, $i(u)$, and firm size under different assumptions about the relative magnitude of unobservable and observable shocks, $\frac{\sigma_U^2}{\sigma_O^2}$.

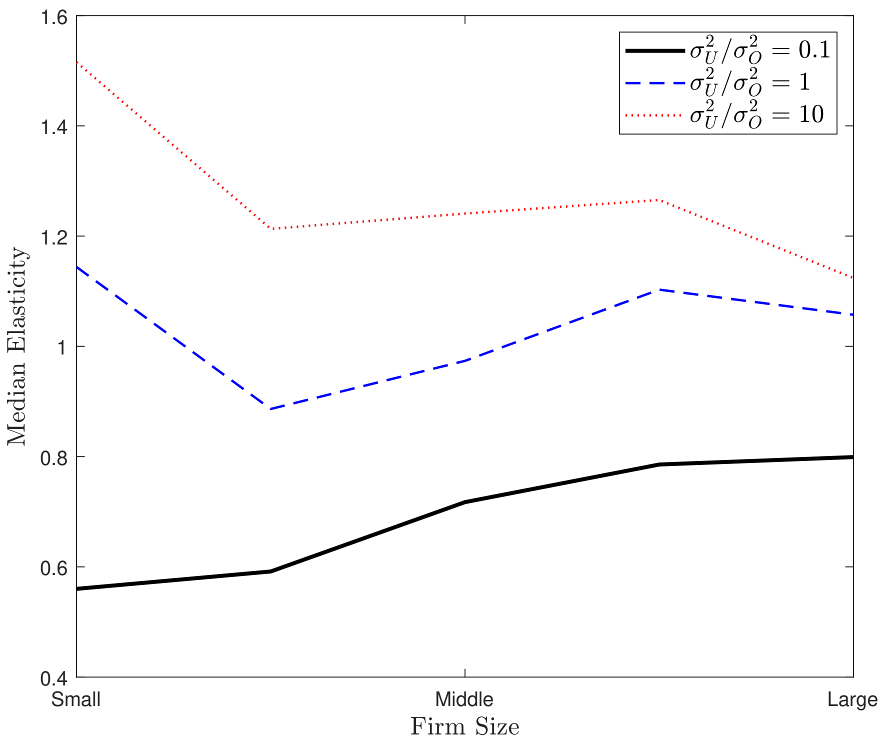
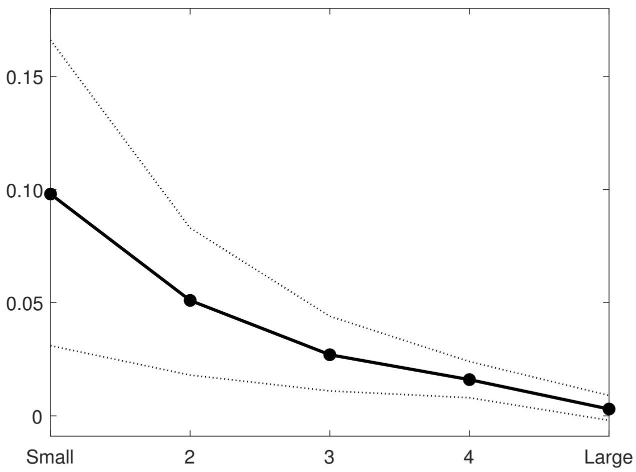
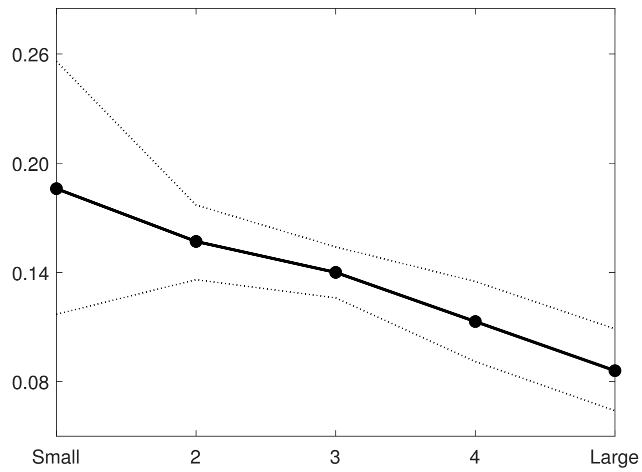


Figure 6. Average Elasticity of Continuation Utility and Firm Size

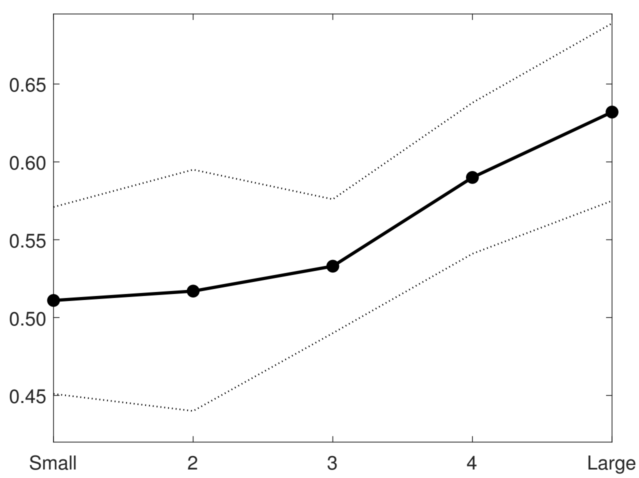
Figure 6 plots the equilibrium relationship between the average elasticity of continuation utility with respect to observable and unobservable shocks (ξ) and firm size under different assumptions about the relative magnitude of unobservable and observable shocks, $\frac{\sigma_U^2}{\sigma_O^2}$.



(a) Growth Rates



(b) Investment Rates



(c) Pay-Performance Sensitivity

Figure 7. Growth, Investment and Pay-Performance Sensitivity

Figure 7 shows variation in average growth and investment rates, and pay-performance sensitivity across five size-sorted portfolios. The dotted lines represent the two-standard error band around the point estimates. Growth rates are measured in logs, PPS is measured in a panel regression of log growth in CEO wealth on log firm return, controlling for firm and time fixed effects. The data are annual, measured in real terms and cover the period from 1992 to 2019.

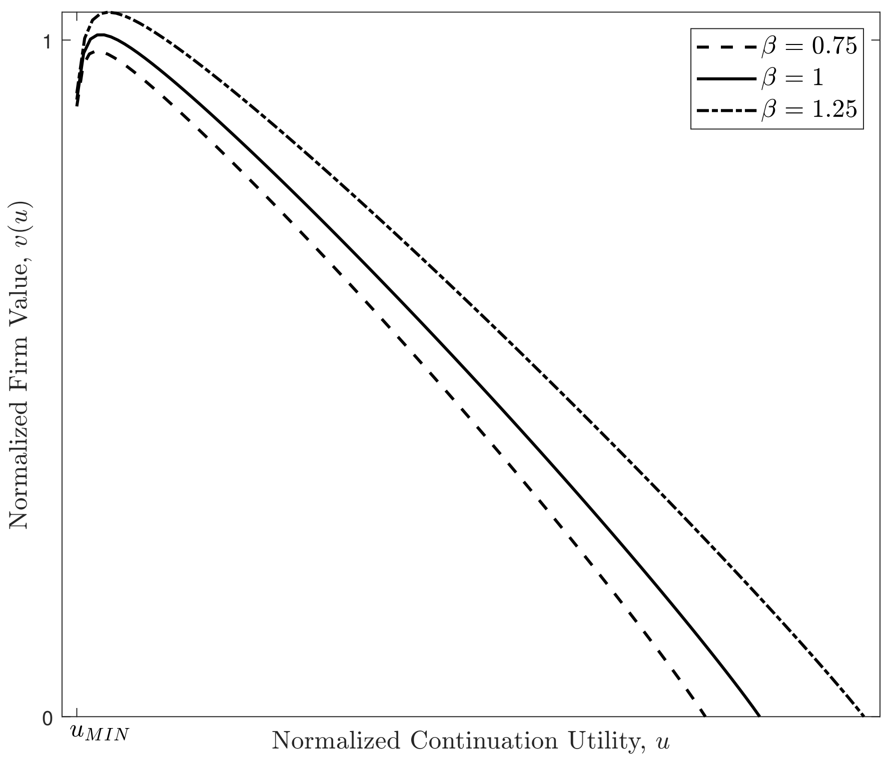


Figure 8. Implications of Aggregate Risk for Normalized Value Function

Figure 8 plots the normalized value functions for different values of aggregate risk exposure: $\beta = 0.75$ (dashed line), $\beta = 1$ (solid line), and $\beta = 1.25$ (dash-dotted line).

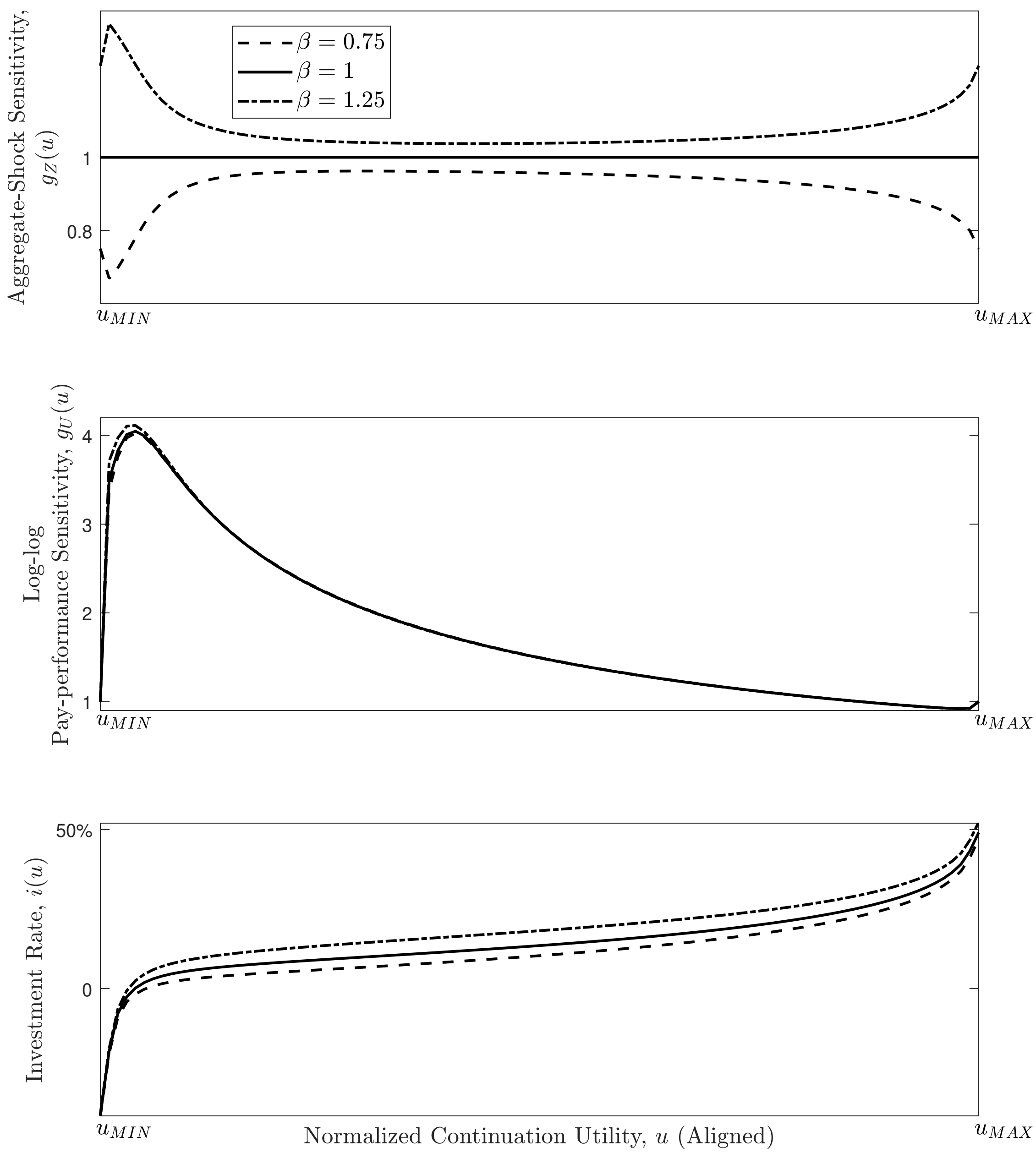


Figure 9. Implications of Aggregate Risk for Policy Functions

Figure 9 plots the policy functions, $g_Z(u)$ (top panel), $g_U(u)$ (middle panel), and $i(u)$ (bottom panel) for different values of aggregate risk exposure: $\beta = 0.75$ (dashed line), $\beta = 1$ (solid line), and $\beta = 1.25$ (dash-dotted line). We normalize u_{MAX} so that the domain of u is the same for all specifications.

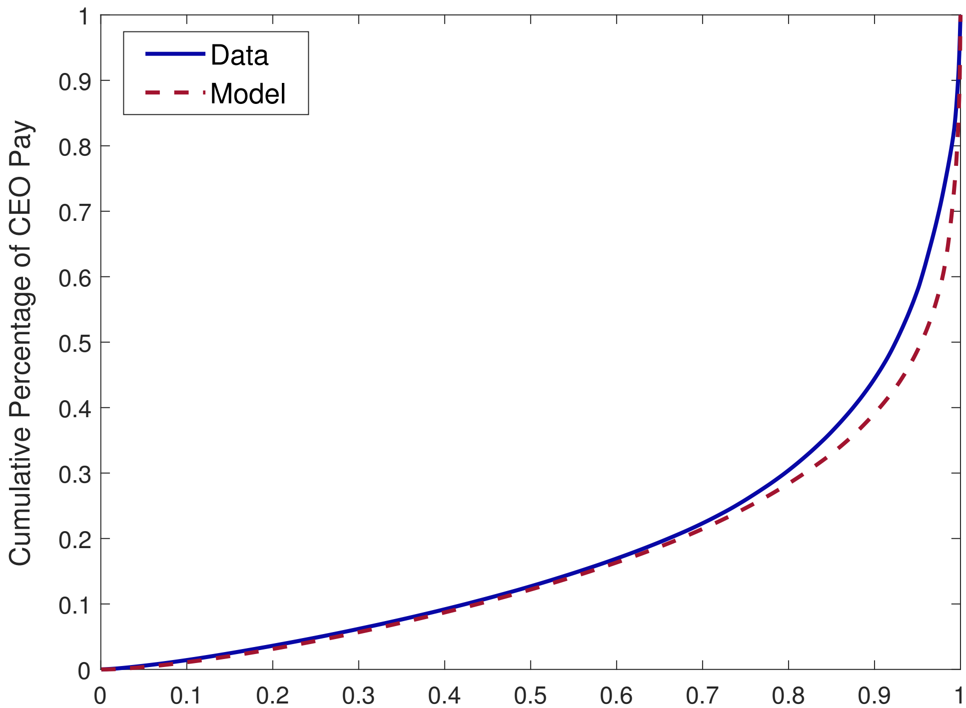


Figure 10. Data and Model-Implied Lorenz Curves

Figure 10 shows data- and model-based Lorenz curves (solid and dashed lines, respectively). The data plot is based on 2018 CEO compensation series, the model plot is based on the estimated baseline model specification.

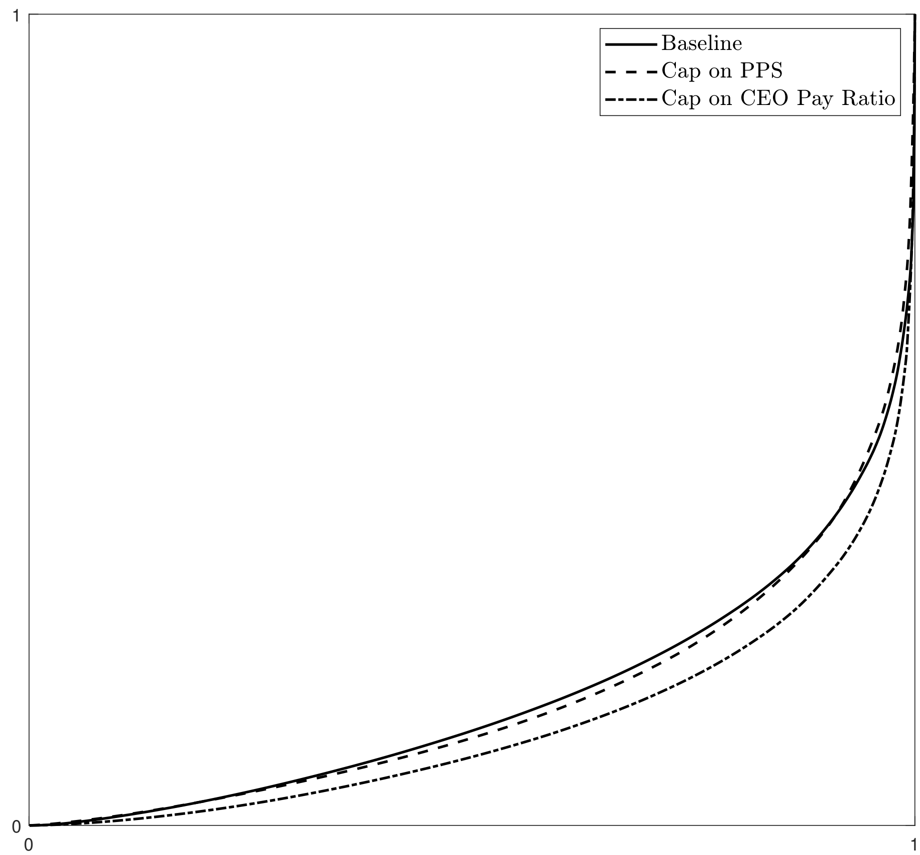


Figure 11. Lorenz Curves for Different Model Specifications

Figure 11 plots the Lorenz curve for our baseline model (solid line), that for the model with an upper bound on PPS (dashed line), and that for the model with an upper bound on CEO pay ratio (dash-dotted line).

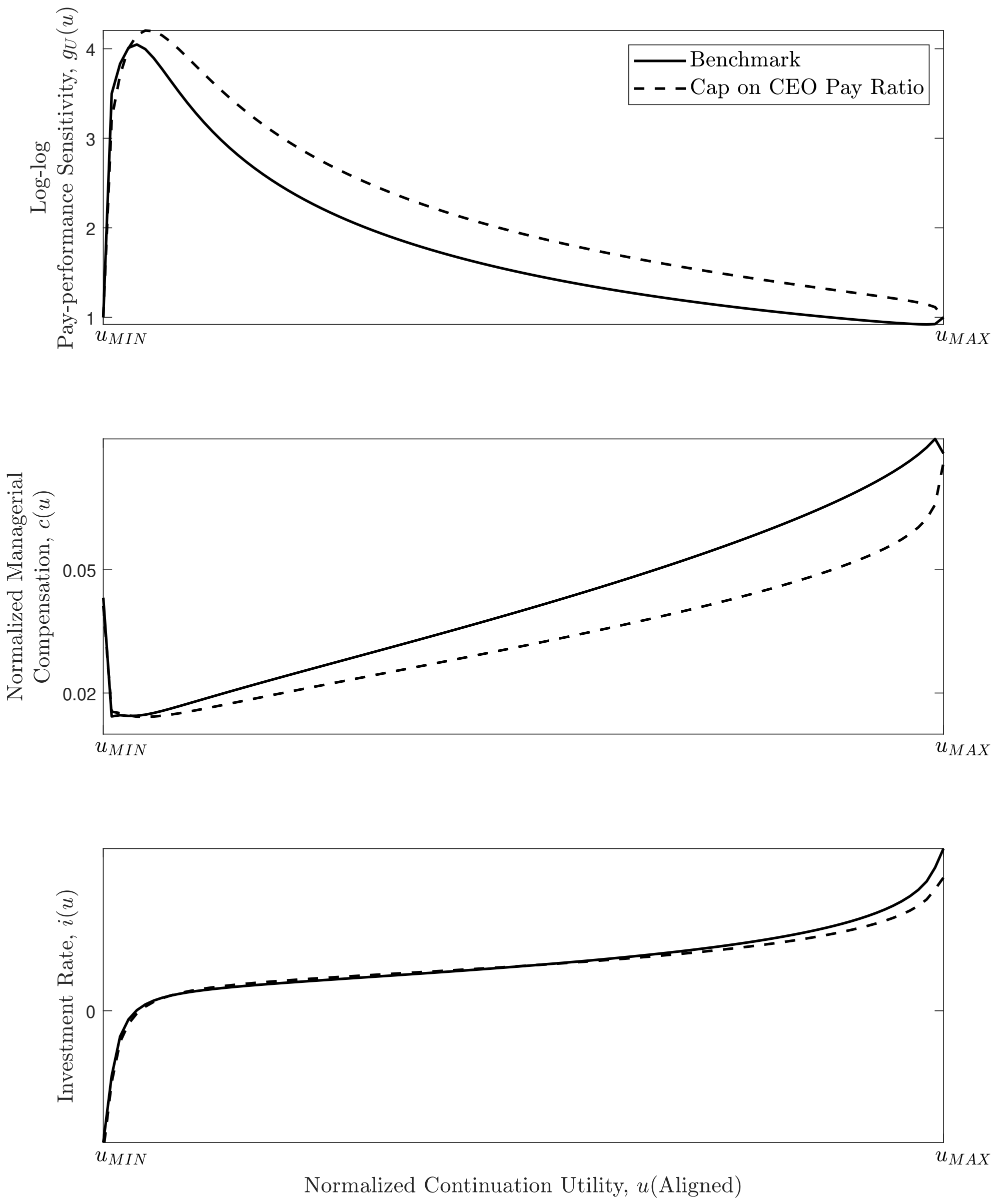


Figure 12. Implications of a Cap on CEO Pay Ratio

Figure 12 plots the policy functions $g_U(u)$ (top panel), $c(u)$ (middle panel), and $i(u)$ (bottom panel) for the benchmark model (solid line) and the corresponding functions for the model with an upper bound on CEO pay ratio (dashed line).

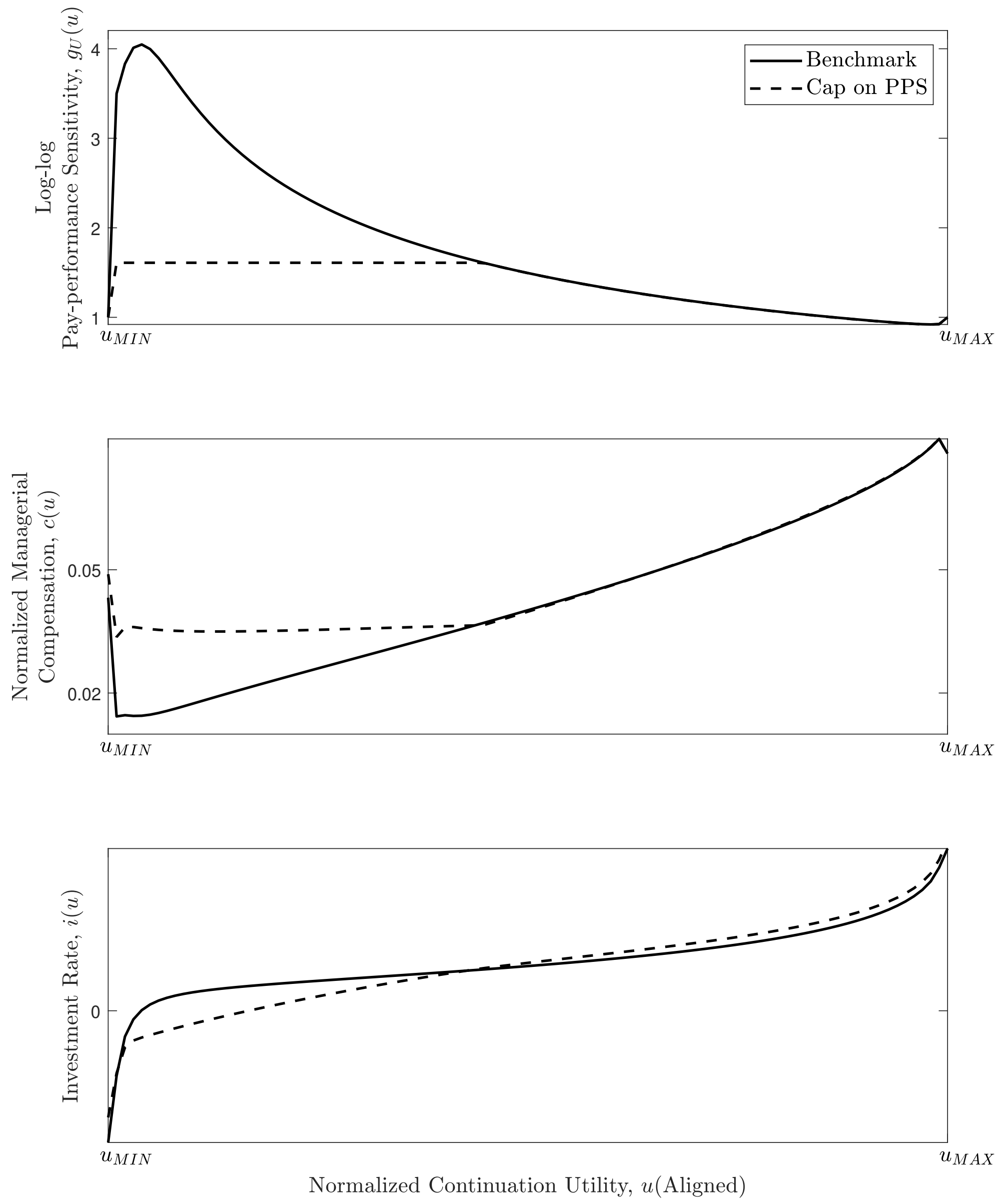


Figure 13. Implications of a Cap on PPS

Figure 13 plots the policy functions $g_U(u)$ (top panel), $c(u)$ (middle panel), and $i(u)$ (bottom panel) for the benchmark model (solid line) and the corresponding functions for the model with an upper bound on log-log PPS with respect to unobservable shocks, g_U (dashed line).