

Moral Hazard, Investment, and Firm Dynamics

Hengjie Ai and Rui Li *

Abstract

We present a dynamic general equilibrium model with heterogeneous firms. Owners of the firms delegate investment decisions to managers whose consumption and investment decisions are their private information. We solve the optimal contracting problem and characterize the implied general equilibrium. Our model provides a quantitative framework to study the cross-sectional distribution and time-series dynamics of firms' investment and dividend payout policies as well as managers' compensation. Risk sharing requires that managers' equity shares decrease with firm sizes. As a result, it is harder to prevent private benefit in larger firms, where managers have lower equity stakes under the optimal contract. In our calibration, small firms invest more, pay fewer dividends, and grow faster than large firms do. Despite the heterogeneity in firms decision rules and the failure of Gibrat's law, we show that the size distribution of firms in our model resembles a power law with a slope coefficient about 1.06, as in the data.

JEL classification: C68, D82, D86, D92, G32.

Keywords: Principal-agent, moral hazard, investment, firm dynamics.

*Hengjie Ai is affiliated with the Fuqua School of Business, Duke University. Rui Li is at the Department of Economics, University of Wisconsin-Madison. We would like to thank Ravi Bansal, Gian Luca Clementi, Dean Corbae, Steven Durlauf, 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 2011 Tel Aviv Finance Conference, the 2012 Conference on Financial Economics and Accounting, the 2012 ASU Sonoran Winter Conference, and the 2012 UBC Winter Finance Conference for useful conversations on the topic. We also thank Song Ma for his excellent research assistance. All errors are our own.

1 Introduction

Small firms invest more ([Gala and Julio \(2011\)](#)), pay fewer dividends ([Fama and French \(2001\)](#)), and grow faster ([Evans \(1987a\)](#), [Evans \(1987b\)](#), [Hall \(1987\)](#)). Our purpose is to understand the underlying economic mechanism for the above phenomenon in a quantitative general equilibrium model.

We emphasize the importance of disciplining the model with the empirical evidence on the power law in the size distribution of firms. In the U.S. data, the distribution of firm sizes resembles closely a power law (or a Pareto distribution) — the number of firms with more than n employees is roughly proportional to $1/n^\xi$, with ξ being close to 1.06 and extremely stable over time. The benchmark model to explain power law is Gibrat’s law, which states that firms’ growth rates are independent of their sizes. The empirical evidence on the dependence of firms’ investment and growth rates on their sizes clearly contradicts the conclusion of Gibrat’s law and presents important challenges to quantitative economic theories.

To account for the above features of the data, we incorporate moral hazard and optimal dynamic contracting into a general equilibrium model with neoclassical production technology. When calibrated to match the standard macroeconomic moments and the volatility of sales at the firm level, our model reproduces the empirical pattern of the dependence of firm’s investment and payout policies on their sizes as well as the power law distribution if unobservable shocks account for a small fraction of the volatility at the firm level.

In our model, shareholders of firms do not have the technology to invest in firms they own and have to delegate their decisions to managers whose consumption and investment decisions are not observable to shareholders. Capital accumulation in firms is determined not only by managers’ investment decisions but also by unobservable and observable idiosyncratic shocks. Because the total output of firms is an increasing function of their capital stock, it serves as a noised signal of managers’ past actions. We assume that both shareholders and managers have identical constant relative risk aversion (CRRA) preferences and derive the optimal dynamic contract that trades off risk sharing against incentive provision. We characterize the general equilibrium induced by the optimal contracts and study the cross-sectional distribution of firms’ characteristics.

Our focus is to understand how the presence of moral hazard affects firms’ investment and

dividend payout policies and their implications on the life-cycle dynamics and the cross-sectional distribution of firms. We find both unobservable and observable shocks to be important in understanding the size-investment and size-dividend payout relationships in the data. In our model, the ratio between a manager's continuation utility and the size of the firm he operates can be interpreted as the manager's equity share in the firm. Qualitatively, the presence of the unobservable shocks gives rise to moral hazard problems and implies that managers have stronger incentive to invest if they have higher equity stakes in the firms they manage. As a result, a firm's investment-to-capital ratio increases in the manager's equity shares.

Quantitatively, the predictions of our model are consistent with the empirical pattern on size-investment and size-dividend payout relationships only when observable shocks account for most of the cross-section dispersion at the firm level. Under the optimal contract, risk sharing with respect to the observable idiosyncratic shocks implies that the manager's equity share in a firm decreases with the amount of favorable observable shocks. Therefore, the managers' equity share is typically high in a small firm and low in a large firm. As a result, small firms invest more, pay fewer dividends, and grow faster due to high managers' equity shares.

In addition, we show that despite the failure of Gibrat's law, the size distribution of firms in our model closely resembles a Pareto distribution with a slope coefficient close to 1.06, as in the data. Firms in our model can get large for one of the two reasons: Some firms are large because they have experienced a sequence of observable positive productivity shocks. Under the optimal contract, the manager's shares in these firms decrease as they grow larger and so do their investment rates. Other firms become large because of their high historical investment rates rather than favorable observable shocks. The latter kind of large firms continue to invest a high fraction of output and therefore grow fast. As a result, although large firms invest less than small ones do on average, a substantial fraction of them continue to grow fast and contribute to the fat tail of the Pareto distribution.

We cast our model in continuous time. Like previous literature, for example, [Sannikov \(2008\)](#), [Williams \(2011\)](#) and [DeMarzo and Sannikov \(2006\)](#), we find that continuous time methods offer a convenient way to characterize the optimal contract.¹ We show that the optimal contract can be characterized as the solution to an ordinary differential equation with known boundary conditions. An additional advantage of our continuous time framework is that it allows a simple characterization

¹In [Ai and Li \(2011\)](#), we show our methodology can be extended to allow for stochastic differential utility.

of the cross-sectional distribution of firms' characteristics. By exploring the homogeneity property of firms' decision rules, we show that the two-dimensional distribution of firm characteristics can be summarized by a one-dimensional measure defined on the space of scaled continuation utilities of the agent. This measure obeys a version of the Kolmogorov forward equation whose solution can be easily calculated numerically.²

Our paper builds on the large literature on contract design and firm dynamics. In the discrete time setting, [Clementi and Hopenhayn \(2006\)](#) and [Clementi *et al.* \(2008\)](#) study optimal dynamic contracts in the presence of moral hazard and its implications on firm dynamics, while [Albuquerque and Hopenhayn \(2004\)](#), [Cooley *et al.* \(2004\)](#) and [Rampini and Viswanathan \(2012\)](#) consider environments where the entrepreneur has limited enforcement. In continuous time setups, [DeMarzo *et al.* \(2009\)](#), [Biais *et al.* \(2010\)](#), [He \(2009\)](#), and [He \(2011\)](#) study optimal dynamic contracting in various settings with moral hazard and investment.

This paper differs from the above literature in several aspects. First, we adopt the quantitative general equilibrium framework. This allows us to draw on the large body of research in dynamic general equilibrium models in macroeconomics to discipline the choices of the preference and technology parameters in our calibration. It also allows us to confront our model to empirical evidence both at the micro and the macro level. For example, our model links the life-cycle dynamics of firms to the cross-section distribution of their characteristics. Because both implications of our model depend on the structural parameters, together, they impose a stronger discipline in quantitative exercises.

Second, our model incorporates some additional features which we believe important in quantitative exercises but are absent from most of the above papers. For example, both the principal and the agent are risk averse in our model, while most of the above mentioned papers focus on risk neutral agents. Although risk neutrality simplifies the contract design problem, it does not capture economic incentives such as intertemporal consumption smoothing and precautionary saving that have been shown to be quantitatively important in understanding many other economic phenomena. In addition, our model features both observable and unobservable shocks. All the above mentioned models that focus on moral hazard problems consider exclusively unobservable shocks. We allow

² [Ai \(2011\)](#) shows that this approach can be applied to a large class of general equilibrium models with heterogeneous agents/firms.

the variation in firms' output to consist of both observable and unobservable shocks and calibrate the relative importance of these shocks to observable moments in the data. Our calibration suggests that most of the volatility of the firms' output comes from observable shocks, although the presence of unobservable shocks is important in understanding the quantitative features of the data.

Third, none of the above papers confront their models with the empirical evidence on the power law distribution of firm size. In fact, many of the above mentioned models assume a decreasing return to scale technology. As a result, the long-run distribution of firm size is likely to have a compact support. In contrast, our paper adopts a constant return to scale technology, and predicts nontrivial firm growth even in the long-run. This feature of our model is crucial in accounting for the Pareto-like size distribution of the firms in the data.

Our paper is also related to the theoretical and empirical literature on firm dynamics and the size distribution of firms. Earlier theoretical works such as [Jovanovic \(1982\)](#), [Hopenhayn \(1992\)](#), and [Cooley and Quadrini \(2001\)](#) focus on the dependence of firm behavior on age and size, but do not confront their models with the empirical evidence on the power law. [Gabaix \(2009\)](#) surveys power laws in economics and finance in general. [Luttmer \(2011\)](#) reviews the recent literature on firm dynamics that generates the power law. Equilibrium models of firm dynamics that emphasize the power law include [Klette and Kortum \(2004\)](#), [Luttmer \(2007\)](#) and [Arkolakis \(2011\)](#), among others. These models typically do not endogenize firm growth as the result of investment and capital accumulation, neither do they micro-found the heterogeneity in firm behavior based on agency frictions. On the empirical side, the relationship between firm size and investment and dividend payout policies has been documented by many researchers. [Mansfield \(1962\)](#) is among the earliest to show that small firms grow faster. [Evans \(1987a\)](#) and [Evans \(1987b\)](#) argue that the size-growth relationship is robust to possible sample selection bias. [Fama and French \(2001\)](#) document that large firms are much more likely to pay dividends than small firms do. The implications of our model are largely consistent with the empirical evidence documented in the above papers.

The rest of the paper is planned as follows. We introduce the setup of the model in Section 2. Section 3 characterizes the optimal dynamic contract for individual firms. Section 4 aggregates firms decisions, studies the cross-sectional distribution of firms' characteristics, and closes the model in general equilibrium. We present our calibration results in Section 5. Section 6 concludes the paper.

2 Setup of the Model

2.1 Preference and Technology

2.1.1 Preferences of Shareholders and Managers

There are a unit measure of shareholders in the economy who can plant trees. The knowledge to grow trees is known only to managers, a special type of agents who arrive at the economy in overlapping generations.³ The shareholders, therefore have to delegate the investment decisions to managers. To keep our language consistent with the principal-agent literature, we will use the terminology shareholders and principals, managers and agents interchangeably.

Time is continuous and infinite. Shareholders and managers have identical preferences represented by expected utility of the form:

$$\int_0^\infty e^{-\beta t} \frac{1}{1-\gamma} C_t^{1-\gamma} dt,$$

where $\gamma > 0$ is the relative risk aversion coefficient, $\beta > 0$ is the discount rate and C_t is the consumption at time t . Shareholders are also endowed with one unit of labor which is supplied inelastically.

2.1.2 Production and Investment Technology

A tree is a technology to combine labor and tree-specific capital to produce consumption goods, and to accumulate tree-specific capital over time. At any point of time, shareholders can plant trees. Trees are indexed by j and each tree is associated with a certain amount of tree-specific capital. We use $Y_{j,t}$, $K_{j,t}$, and $N_{j,t}$ to denote the total output produced, the amount of capital stock and labor employed by tree j at time t respectively. A tree combines capital and labor to produce consumption goods via the standard Cobb-Douglas technology:

$$Y_{j,t} = z K_{j,t}^\alpha N_{j,t}^{1-\alpha},$$

³The overlapping generation aspect of our model follows the continuous time "perpetual youth" model of [Blanchard \(1985\)](#).

where z is an economy-wide common productivity parameter. We focus on the case in which z is a constant.

Let W_t denote the real wage at time t , the total operating profit of tree j at time t is

$$\Pi(K_{j,t}) = \max_{N_{j,t}} \{zK_{j,t}^\alpha N_{j,t}^{1-\alpha} - W_t N_{j,t}\}. \quad (1)$$

Tree-specific capital can be accumulated according to the following investment technology:

$$dK_{j,t} = K_{j,t} [-\delta dt + \sigma^T dB_{j,t}] + I_{j,t} dt, \quad (2)$$

where $I_{j,t}$ is the total investment chosen by the manager of the tree. δ is the depreciation rate of capital common among all trees. The term $B_{j,t}$ is a 2×1 vector of standard Brownian motions independent across firms, and

$$\sigma^T dB_j = \sigma_u dB_{u,j} + \sigma_o dB_{o,j}.$$

In the above expression, $\sigma_u, \sigma_o > 0$ are constants. The Brownian motion $B_{u,j}$ is unobservable to all except the manager who operates the tree, and the Brownian motion $B_{o,j}$ is common knowledge in the economy. $B_{u,j}$ and $B_{o,j}$ are independent.

2.1.3 Information and Managerial Compensation

At any point of time t , given the total amount of capital stock $K_{j,t}$, which is observable to all, the manager makes observable decisions on labor hiring $N_{j,t}$. After the wage bill $W_t N_{j,t}$ is paid, the manager hand in the total operating profit, $\Pi(K_{j,t})$, to shareholders. The total operating profit is divided among dividend payment to shareholders, $D_{j,t}$, compensation to the manager, $C_{j,t}$, investment in the firm, and a cost of capital adjustment of the form $H(I_{j,t}, K_{j,t})$:

$$D_{j,t} + C_{j,t} + I_{j,t} + H(I_{j,t}, K_{j,t}) = \Pi(K_{j,t}). \quad (3)$$

We assume that the dividend has to be non-negative and cannot exceed the total operating profit:

$$0 \leq D_{j,t} \leq \Pi(K_{j,t}). \quad (4)$$

The consumption and investment decisions of the manager are observable only to himself. Given operating profit $\Pi(K_{j,t})$ and dividend policy $D_{j,t}$, equation (3) imposes a constraint on manager's consumption and investment choices:

$$C_{j,t} + I_{j,t} + H(I_{j,t}, K_{j,t}) = \Pi(K_{j,t}) - D_{j,t}.$$

We allow for a constant return to scale adjustment cost:

$$H(I, K) = h\left(\frac{I}{K}\right) \cdot K. \quad (5)$$

We use quadratic adjustment cost: $h(i) = \frac{1}{2}\bar{h}\left(\frac{I}{K} - i^*\right)^2$ with $\frac{1}{2i^*} > \bar{h} \geq 0$. The special case with $\bar{h} = 0$ corresponds to zero adjustment cost. The assumption $\bar{h} < \frac{1}{2i^*}$ guarantees that the agent can always obtain a positive amount of consumption by setting $I_j(t)$ negative regardless of shareholder's policy. We assume that there is a minimum level of investment-to-capital ratio, that is,

$$\frac{I_{j,t}}{K_{j,t}} \geq -B,$$

for some large real number $B > 0$. This is merely a technical assumption ensuring that the manager's optimization problem has a compact choice set and therefore a well-defined solution. We will chose B to be large enough so that the above constraint never binds in equilibrium.

2.1.4 Entry and Exit of Firms

A unit measure of managers arrive at the economy at each point of time. Upon arrival, a manager is endowed with the technology to operate a tree, and an outside option that delivers a reservation utility U_0 ⁴. The reservation utility of newly arrived managers is distributed according to a probability measure Λ with support $[U_L, U_H]$.

At each point of time, the principal offers a contract to every newly arrived manager that specifies the initial size of the tree to be delegated to the manager and future payment to the manager as a function of the history of all past realizations of observables and actions. If the contract is accepted,

⁴The reservation utility can be endogenized by assuming that managers have access to an alternative technology that delivers exactly the life time utility U_0 . We do not spell out the details here as in equilibrium, this option is never chosen.

the principal will plant the tree and delegate the tree to the manager until he exits the economy. Planting a tree with initial size K_0 costs exactly K_0 amount of consumption good.

Optimality requires that shareholders offer contracts to managers that deliver exactly their reservation utility upon arrival. We assume full commitment on both sides. In this case all managers will accept their contracts and work for the trees until they exit the economy. Given $U_0 \in [U_L, U_H]$, shareholders will choose the initial size of the tree optimally. Let $K_0(U_0)$ denote the optimal initial size of the tree delegated to a manager with reservation utility U_0 . The total amount of resources used to plan new trees at each moment is therefore

$$\int_{[U_L, U_H]} K_0(U) \Lambda(dU).$$

We assume that the market is complete, so that shareholders can fully diversify any idiosyncratic shocks. However, because of the moral hazard problem, the only way that managers can credibly participate in the credit market is through the compensation contracts that tie their consumption to the performance of the tree that he operates.

Each tree receives a blight disease shock with an exogenous Poisson rate $\kappa > 0$ per unit of time. Once the shock hits, the tree stops producing fruits that can be consumed by shareholders and then exits the economy. Let τ denote the stopping time at which the tree is hit by the shock. The manager of the tree also exits the economy at time τ with a terminal utility $T(K_\tau)$. We assume

$$T(K) = \frac{(u_T K)^{1-\gamma}}{1-\gamma}, \quad u_T > 0$$

to be homogeneous of degree of $1-\gamma$ to keep the homogeneity of manager's preference. $u_T > 0$ is a scaling parameter. Because a unit measure of trees are planted per unit of time, and trees exit the economy at rate κ , the total measure of trees in the economy in steady state is $\frac{1}{\kappa}$.

We think of each tree as a firm and use the terminology trees and firms interchangeably from now on. In our setup, each firm is identified by the firm-specific technology and the contractual relationship between the shareholders and the manager of the firm.

2.2 Shareholders' and Managers' Optimization Problem

2.2.1 Feasibility of Plans

Tree-manager pairs are indexed by j . We use $\tau_{j,0}$ to denote the calendar time at which manager j arrives at the economy, and we use τ_j to denote the time at which tree j is hit by the blight disease and exits the economy. For each tree j , shareholders choose a dividend policy $D_{j,t}$, investment policy $I_{j,t}$ and consumption of the manager, $C_{j,t}$ for all $\tau_{j,0} \leq t < \tau_j$. We use a superscript to denote the history of the realizations of a stochastic process up to time t . For example, K_j^t refers to the realizations of the firm- j specific capital from time $\tau_{j,0}$ to time t , $\{K_{j,s}\}_{s \in [\tau_{j,0}, t]}$. The consumption, investment, and dividend payment policies are functions of observables, that is, the history, $\{K_j^t, B_{j,o}^t\}$, denoted by H_j^t . For $s \geq t$, we use the notation $H_j^s \geq H_j^t$ to denote a history that follows H_j^t . To simplify notation, we suppress the firm subscript j if no confusion arises. Formally, a plan is a triple, $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$ that specifies managerial consumption, investment, and dividend payout as a function of the history, $H^t = \{K^t, B_o^t\}$.

Given a plan, the manager chooses his consumption and investment policies $\{\tilde{C}_t, \tilde{I}_t\}_{t \in [\tau_0, \tau]}$ conditioning on his own information, which contains the history of the Brownian motions, B_u^t and B_o^t , as well as that of the manager's own past actions. We define a strategy of the manager to be a consumption and investment policy, $\{\tilde{C}_t, \tilde{I}_t\}_{t \in [\tau_0, \tau]}$, adapted to the filtration generated by the Brownian motions, B_u and B_o . Because dividend payment is observable to the principals, a strategy, $\{\tilde{C}_t, \tilde{I}_t\}_{t \in [\tau_0, \tau]}$, is feasible given a plan, $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$, if it is adapted to the manager's information filtration and satisfies

$$\tilde{C}_t + \tilde{I}_t + H(\tilde{I}_t, K_t) = \Pi(K_t) - D_t(H^t) = C_t(H^t) + I_t(H^t) + H(I_t(H^t), K_t), \text{ for all } t \text{ and } H^t. \quad (6)$$

That is, given a plan, feasibility of manager's strategy requires it to be consistent with the dividend policy prescribed by the plan. Since consumption and investment policy are unobservable to the principals, a manager can choose an alternative strategy other than that prescribed by the plan in order to maximize his utility without being detected.

Given a policy $\{\tilde{C}_s, \tilde{I}_s\}_{s \in [\tau_0, \tau]}$, a manager's continuation utility at time $t \geq \tau_0$ is computed as

$$U_t \left(\{\tilde{C}_s\}_{s \in [t, \tau]} \right) \equiv E_t \left[\beta \int_t^\tau e^{-\beta(s-t)} \frac{\tilde{C}_s^{1-\gamma}}{1-\gamma} ds + e^{-\beta\tau} T(K_\tau) \right],$$

where the conditional expectation is taken with respect to manager's information set at time t . Formally, it is conditioned on the filtration generated by the two dimensional Brownian motion $\{B_{u,s}, B_{o,s}\}_{s \in [\tau_0, t]}$.

Given a plan $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$, at time $t \geq \tau_0$, the manager chooses a continuation strategy $\{\tilde{C}_s, \tilde{I}_s\}_{s \in [t, \tau]}$ from the set of all feasible strategies, to maximize $U_t \left(\{\tilde{C}_s\}_{s \in [t, \tau]} \right)$. A plan, $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$, is *incentive compatible* if and only if, for any H^t , $C_t(H^t)$ and $I_t(H^t)$ are the optimal choice of the manager. Equivalently, a plan is incentive compatible if and only if for any t and any history up to time t .

$$U_t \left(\{C_s(H^s), I_s(H^s)\}_{s \in [t, \tau]} \right) \geq U_t \left(\{\tilde{C}_s, \tilde{I}_s\}_{s \in [t, \tau]} \right), \quad (7)$$

for any alternative feasible continuation strategy, $\{\tilde{C}_s, \tilde{I}_s\}_{s \in [t, \tau]}$ that satisfies condition (6). Following Myerson (1997)'s language, we will call (7) the obedience constraint.

A plan $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$ is said to be *feasible* with promised utility U if it satisfies constraints (3), (4), the incentive compatibility condition in (7) and participation constraint

$$U_{\tau_0} \left(\{C_t, I_t\}_{t \in [\tau_0, \tau]} \right) = U.$$

The objective of the shareholders is to chose among the set of feasible plans to maximize her profit, which we turn now to.

2.2.2 Shareholder Value Maximization

We focus on the steady-state of the economy where the aggregate consumption of shareholders is constant. This implies that the risk-free interest rate, r , equals the discount rate of the shareholders:

$$r = \beta.$$

Shareholders choose a plan, $\{C_t(H^t), I_t(H^t), D_t(H^t)\}_{t \in [\tau_0, \tau]}$, to maximize the present value of firms:

$$E \left[\int_{\tau_0}^{\tau} e^{-r(t-\tau_0)} D_t dt \right] \quad (8)$$

subject to the feasibility constraint.

2.2.3 Recursive Formulation

In this subsection, we provide a recursive formulation to describe the allocation for each firm. The history of realizations of observables and past actions can be summarized by two state variables (K, U) . Let \mathcal{K} denote the space of possible realizations of capital, and \mathcal{U} denote the space of promised utilities. Using the language of [Atkeson and Lucas \(1992\)](#), we define an allocation rule to be a collection of functions,

$$C(K, U), I(K, U), D(K, U), N(K, U), G_u(K, U) \text{ and } G_o(K, U)$$

that maps the state space $\mathcal{K} \times \mathcal{U}$ into the real line. Given an allocation rule, we can recover allocation recursively. First, for a given $t \in [\tau_0, \tau)$, determine manager consumption, investment, dividend payout, and amount of labor hired given state variables (K_t, U_t) by

$$C_t = C(K_t, U_t); \quad I_t = I(K_t, U_t); \quad D_t = D(K_t, U_t), \quad N_t = N(K_t, U_t).$$

Second, use the law of motion of state variables, along with the allocation rule to construct future state variables for all possible realizations of shocks:

$$\begin{aligned} dU_t = & \left(\beta \left(U_t - \frac{1}{1-\gamma} C(K_t, U_t)^{1-\gamma} \right) - \kappa(T(K_t) - U_t) \right) dt \\ & + G_u(K_t, U_t) \sigma_u dB_{u,t} + G_o(K_t, U_t) \sigma_o dB_{o,t} \end{aligned} \quad (9)$$

and

$$dK_t = K_t \left[\left(\frac{I(K_t, U_t)}{K_t} - \delta \right) dt + \sigma_u dB_{u,t} + \sigma_o dB_{o,t} \right]. \quad (10)$$

Given the initial condition $(K_0(U_{\tau_0}), U_{\tau_0})$, the above procedure fully describes a plan. Without loss of generality, optimal plans can be constructed by using allocation rules.⁵

We provide a formal derivation of the law of motion of continuation utility in (9) in Appendix A.2. Intuitively, the policy functions $G_u(K, U)$ and $G_o(K, U)$ describe the rules of assigning continuation utilities based on the realizations of the Brownian shocks. At each point of 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, $C(K_t, U_t)$, a rate of change of continuation utility with respect to unobservable shocks, $dB_{u,t}$, $G_u(K_t, U_t)$, and a rate of change of continuation utility with respect to observable shocks, $dB_{o,t}$, $G_o(K_t, U_t)$. A continuation contract from time t on delivers the utility level U_t as long as (9) is satisfied for all $t < \tau$ and

$$U_\tau = T(K_\tau). \quad (11)$$

In this formulation, the constraints (3) and (4) are written as:

$$C(K, U) + I(K, U) + H(I, K) + D(K, U) = \Pi(K). \quad (12)$$

and

$$0 \leq D(K, U) \leq \Pi(K), \quad (13)$$

respectively.

Given an allocation rule, incentive compatibility can be reduced to the optimality condition of the manager as summarized by the following lemma:

Lemma 1 (Incentive Compatibility) *A plan constructed from the allocation rule, $C(K, U)$, $I(K, U)$, $D(K, U)$, $N(K, U)$, $G_u(K, U)$, $G_o(K, U)$ satisfies the obedience constraint (7) if and only if for all (K, U) ,*

$$\begin{aligned} (C(K, U), I(K, U)) &\in \arg \max_{C, I} \beta \frac{1}{1-\gamma} C^{1-\gamma} + \frac{G_u(K, U)}{K} I \\ \text{subject to } C + I + H(I, K) &= C(K, U) + I(K, U) + H(I(K, U), K) \end{aligned} \quad (14)$$

⁵Without loss of generality, all optimal payoffs can be achieved by allocations generated by recursive allocation rules described here. For formal justification of this result, see Abreu *et al.* (1990), Spear and Srivastava (1987).

In particular, whenever $(C(K, U), I(K, U))$ is interior,⁶ the above is equivalent to

$$\frac{G_u(K, U)}{K} = \beta C(K, U)^{-\gamma} \cdot [1 + H_I(I(K, U), K)]. \quad (15)$$

Proof. See Appendix A.3. ■

Condition (15) highlights the intertemporal choice problem of the manager. The term $\beta C(K, U)^{-\gamma}$ is the marginal utility of consumption of the manager. $1 + H_I(I(K, U), K)$ is the cost of investment goods in terms of current period consumption numeraire. Therefore the left-hand side of equation (15) is the marginal cost of investment in utility terms. Because shareholders do not observe the Brownian motion dB_u , they assign continuation utilities according to the realized capital stock K_t . Therefore, from the manager's perspective, $\frac{G_u(K, U)}{K}$ measures the increase in continuation utility for one additional unit of capital. Obedience requires that the allocation rule for manager's consumption-investment decisions must be optimal from the manager's perspective. Equation (15) can be interpreted as the Q-theory relation from the perspective of the manager: optimality requires that the marginal benefit of an additional unit of investment equals its marginal cost. In this formulation, an allocation rule is said to be feasible if it satisfies conditions (12), (13), and is incentive compatible.

2.3 Definition of Equilibrium

We consider the competitive equilibrium of the economy where the claims to equity of all firms are traded. We focus on the steady-state in which all aggregate quantities are constant. Thanks to the recursive formulation discussed above, firms can be identified by the state variables (K, U) which capture the heterogeneity of the trees. Let $\Phi(U, K)$ denote the density of the distribution of firms' type in steady-state. The equilibrium allocation consists of the choice of dividend payment, managerial compensation, investment, continuation utility assignment rules, and amount of labor hired for firms of all types:

$$\{D(U, K), C(U, K), I(U, K), G_u(U, K), G_o(U, K), N(U, K)\}_{(K, U) \in \mathcal{K} \times \mathcal{U}},$$

⁶The definition of interior solution of consumption-investment plan is given in Appendix A.3.

the initial capital of newly built firms, K_0 , and the total amount of consumption of shareholders, \mathbf{C}^P . Equilibrium prices include wage rate W , the price of equity for firms of each type, $\{V(U, K)\}_{U, K}$, and equilibrium interest rate r . A competitive equilibrium is a list of equilibrium allocations and prices such that

1. Maximization of operating profit on the product market: given the equilibrium wage, firms of all types choose $N(U, K)$ to maximize operating profit as in (1).
2. Shareholder value maximization: given the equilibrium interest rate, for firms of all types, the allocation $\{D(U, K), C(U, K), I(U, K), G_u(U, K), G_o(U, K)\}$ maximizes the value of equity defined in (8) among all feasible allocations.
3. Intertemporal maximization of shareholders: equilibrium interest rate is constant with shareholders' intertemporal optimization, that is, $r = \beta$ in steady-state.
4. Labor market clearing: the total amounts of labor hired by all firms sum up to the total labor endowment, 1:

$$\int \int N(U, K) \Phi(U, K) dU dK = 1 \quad (16)$$

3 The Optimal Contract

3.1 Homogeneity of Value Function and Decision Rules

In general, the optimal contract depends on two state variables, (K, U) . The constant return to scale technology together with the homogeneity of preferences allow us to obtain a key simplification of the contracting problem: all decision rules are homogeneous of degree 1 with respect to K . We first show that given equilibrium prices, firms' operating profit is linear in K . This result is due to Hayashi (1982)' and is summarized by the following lemma. Intuitively, because the labor market is perfectly competitive, the optimal allocation of labor equalizes the marginal product of capital across all firms.

Lemma 2 (Linearity of the Operating Profit Function) *The operating profit function $\Pi(K)$ is linear. In steady-state, the operating profit function is given by;*

$$\Pi(K) = AK,$$

where the constant A is the marginal product of capital:

$$A = \alpha z \mathbf{K}^{\alpha-1}, \tag{18}$$

and the bold faced letter \mathbf{K} stands for the total capital stock in the economy in steady state.

Proof. See Appendix A.1.1. ■

Because the operating profit function is linear in K and the manager's utility function is homogeneous of degree $1 - \gamma$, the value function of the principal, $V(K, U)$ satisfies the following homogeneity property:

$$V(K, U) \equiv v \left(\frac{((1 - \gamma)U)^{\frac{1}{1-\gamma}}}{K} \right) \cdot K \equiv v(u) K \tag{19}$$

for some $v : \mathbb{R} \mapsto \mathbb{R}$. We define

$$u = \frac{((1 - \gamma)U)^{\frac{1}{1-\gamma}}}{K} \tag{20}$$

to be the normalized utility, and call $v(u)$ the normalized value function. Note also, $v(u)$ is the "average Q" of the firm, i.e. the ratio of the total value of the firm divided by its total capital stock.

We also normalize the policies and write:

$$c(u) = \frac{C(K, U)}{K}; \quad i(u) = \frac{I(K, U)}{K}; \quad d(u) = \frac{D(K, U)}{K}; \quad n(u) = \frac{N(K, U)}{K}$$

$$g_u(u) = \frac{G_u(K, U)}{(1-\gamma)U}; \quad g_o(u) = \frac{G_o(K, U)}{(1-\gamma)U}.$$

Using the above notation, condition (14) can be written as

$$(c(u), i(u)) \in \arg \max_{\tilde{c}, \tilde{i}} \frac{\beta}{1-\gamma} \tilde{c}^{1-\gamma} + \tilde{i} g_u u^{1-\gamma} \quad (21)$$

$$\text{subject to} \quad : \quad c + i + h(i) = c(u) + i(u) + h(i), \quad (22)$$

and whenever $(c(u), i(u))$ is interior, the above condition can be reduced to:

$$g_u = \beta c^{-\gamma} u^{\gamma-1} [1 + h'(i)].$$

Shareholders' value maximization implies that the normalized value function $v(u)$ must satisfy an optimality condition represented by the Hamilton-Jacobi-Bellman (HJB) equation. This is summarized in the following proposition.

Proposition 1 *Suppose $v(\cdot)$ is the normalized value function defined by (19), then $v(\cdot)$ is the solution of the following HJB equation*

$$0 = \max_{c, i, d, g_u, g_o} \left\{ \begin{aligned} & d + [i - r - \kappa - \delta] v(u) \\ & + uv'(u) \left[\frac{\beta}{1-\gamma} \left(1 - \left(\frac{c}{u} \right)^{1-\gamma} \right) - \frac{\kappa}{1-\gamma} \left(\left(\frac{u_T}{u} \right)^{1-\gamma} - 1 \right) - (i - \delta) + \gamma (g_u^2 \sigma_u^2 + g_o^2 \sigma_o^2) \right] \\ & + \frac{1}{2} u^2 v''(u) \left[(g_u - 1)^2 \sigma_u^2 + (g_o - 1)^2 \sigma_o^2 \right] \end{aligned} \right\} \quad (23)$$

subject to (21) and

$$0 \leq c + i + h(i) + d \leq A \quad (24)$$

Proof. See Appendix A.5. ■

The normalized value function is fully determined by specifying appropriate boundary conditions. We provide the details of the boundary conditions in Appendix A.4. Figure 1 depicts the

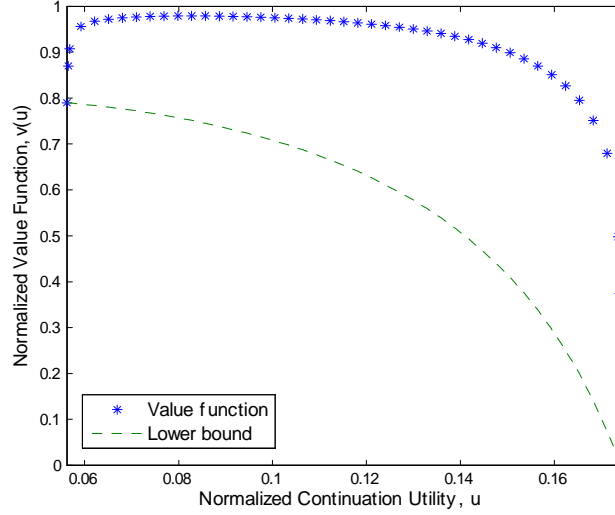


Figure 1: Normalized Value Function and Lower Bound.

normalized value function of the firm, where all parameters are the same as those used in the calibration in Section 5. The right boundary point of normalized utility, u_H corresponds to the allocation where all operating profit is given to the manager's at each moment, or equivalently, the manager owns the entire firm by himself. Here $v(u_L) = 0$ because the firm never pays dividend to shareholders. The left boundary u_L corresponds to the allocation where all operating profit is distributed as dividend to shareholders at each moment. In this case, the manager chooses his consumption optimally by setting investment negative. u_H is the highest and u_L is the lowest possible utility that the shareholder can deliver to the manager. The normalized utility can also be interpreted as a measure of manager's equity share in the firm. The boundary point u_H can be interpreted as manager owning 100% of the equity of the firm, and u_L is the lowest equity share of the manager under the contract.

3.2 Investment Rate and Managers' Equity Share

We first show the marginal benefit of investment (marginal Q) increases with the normalized utility, u . As a result, investment is typically high in firms with high manger equity share.

The intuition for the above result can be explained by the managers' incentive compatibility condition. As we show in Lemma 2, obedience implies that investment specified under the optimal

contract must be optimal for the manager. Investment tends to be higher in firms where managers have larger equity shares because shareholders' and managers' incentives are better aligned with each other.

The monotonicity of marginal Q with respect to u is also consistent with shareholder's optimality condition. Firm value maximization requires investment policy to be optimal from the shareholder's perspective as well. Due to risk-sharing motives, shareholders prefer higher investment rates in firms with larger manager equity share. Intuitively, the marginal cost of utility provision increases with manager's equity share. At u_H , the manager owns the entire firm, therefore the only way to deliver utility is to give all operating profit to the manager. This is a very costly way to provide utility, as there is no risk sharing and the manager's consumption stream is very volatile. At low levels of u , shareholders can achieve a maximum amount of risk sharing by conditioning future assignment of utility on realizations of unexpected shocks. Consequently, it is optimal for the shareholder to invest more at high levels of u as it lowers future levels u by increasing total capital stock K .

We formalize the above result in the following proposition.

Lemma 3 *Suppose $V(K, U)$ is strictly concave in K . $V_K(K, U)$ is strictly increasing in u over (u_L, u_H) .*

Proof. By equation (42) in Appendix A.5, $\frac{d}{du}V_K(K, U) = -v''(u)$. By (43) and concavity of $V(K, U)$ in K , we have the result. ■

In the absence of adjustment cost, the marginal cost of investment is always one. Investment will be positive whenever the marginal benefit of investment, $V_K(K, U)$ exceeds one, and negative if $V_K(K, U)$ is lower than one. This is a direct consequence of the above proposition, which we summarize in the following corollary.

Corollary 1 *Suppose $V(K, U)$ is strictly concave in K . Then there exists a $u_{SWITCH} \in (u_L, u_H)$ such that the marginal benefit of investment at u_{SWITCH} ,*

$$v(u_{SWITCH}) - u_{SWITCH}v'(u_{SWITCH}) = 1.$$

Furthermore, $\forall u > u_{SWITCH}$, $c(u) + i(u) = A$ and $i(u) > 0$. $\forall u < u_{SWITCH}$, $c(u) + i(u) = 0$ and $i(u) < 0$.

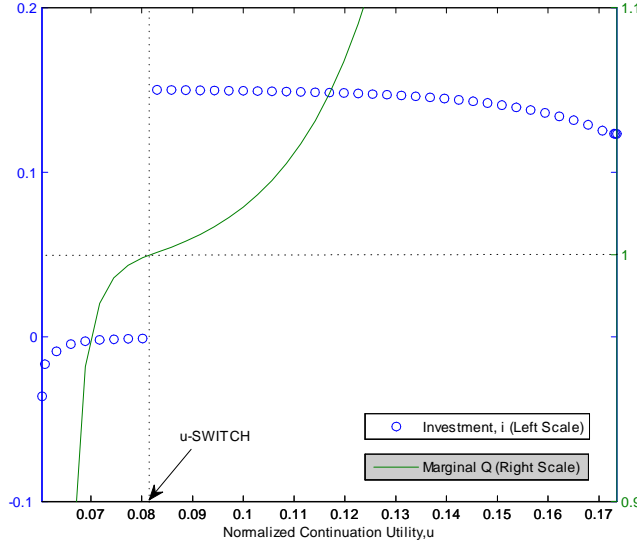


Figure 2: Marginal Q and Investment without Adjustment Cost.

Proof. See Appendix A.2. ■

Corollary 1 characterizes the optimal investment and dividend payout policy in the case of no adjustment cost, which are illustrated in Figure 2 and 3. Figure 2 shows marginal Q (right scale) and optimal investment (left scale) as functions of normalized utility u . Figure 3 depicts manager consumption (squared line), investment (circled line), and dividend payout (dotted line) normalized by total capital stock K . As shown in Figure 3, marginal Q increases with u . The marginal benefit of investment is exactly one at u_{SWITCH} . To the right of u_{SWITCH} , managers' equity share is high, and the marginal Q is higher than marginal cost. In this case, it is optimal for the shareholder to minimize dividend payment and invest as much as possible. As a result, $d(u) = 0$ and $c(u) + i(u) = A$ as shown in Figure 3. Because of the Inada condition, consumption cannot be negative; therefore $0 < i(u) < A$, as shown in Figure 2. In Figure 3, we see that to the right of u_{SWITCH} , investment rate decreases and manager consumption increases slowly. Although marginal Q increases with u , maximum investment level is bounded by the constraint $c(u) + i(u) \leq A$. In fact because increase in normalized utility must be achieved by increase in consumption, investment in fact decreases.

To the left of u_{SWITCH} , the marginal benefit of investment is strictly less than its marginal cost. As a result, the shareholder prefers to disinvest as much as possible and distribute all operating profit

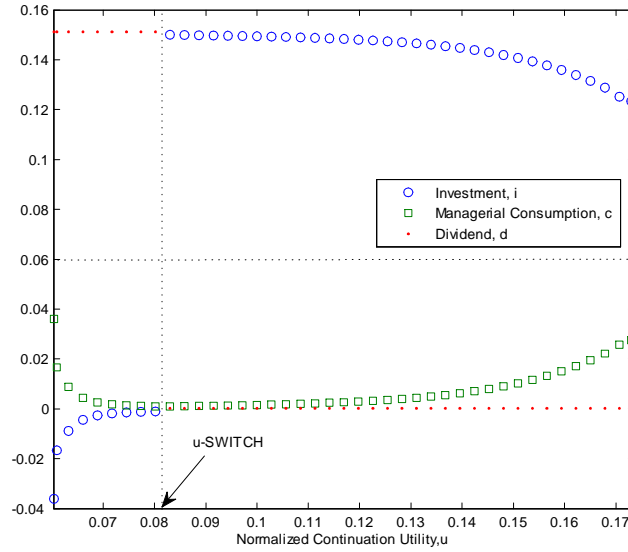


Figure 3: Managerial Consumption, Investment and Dividend Payout without Adjustment Cost.

as dividend. In this case, $d(u) = A$ and $c(u) + i(u) = 0$ as show in Figure 3. Since consumption must be positive, $i(u) < 0$ in this region. Here investment is increasing in the manager's equity share and increase in promised utility is achieved by increase in investment in the firm.

Our calibrated model allows for quadratic adjustment cost. In this case, the marginal cost of investment increases gradually and optimal investment becomes a smooth function of u as shown in Figure 4. Figure 5 illustrates manager consumption (squared line), investment (circled line), and dividend payout (dotted line) normalized by total capital stock K for the model with adjustment cost. We use the parameter values calibrated in Section 5 of the paper to plot Figure 4 and 5. Overall, investment rate increases with normalized utility u and dividend decreases with u . Manager's consumption normalized by capital stock is decreasing in u in the left region of the promised utility space. In this case, higher promised utility is realized by reducing the manager's current consumption, increasing his future consumption and by increasing investment. Due to the constraint $c + i + h(i) + d \leq A$, investment rate cannot increase indefinitely. Eventually, higher promised utility has to be realized by providing more current consumption and manager's consumption becomes increasing in u .

Variation in marginal Q induced by concavity of the firm's value function is the main determinant of firm's investment decision. This aspect of our model is similar to the neoclassical model. The

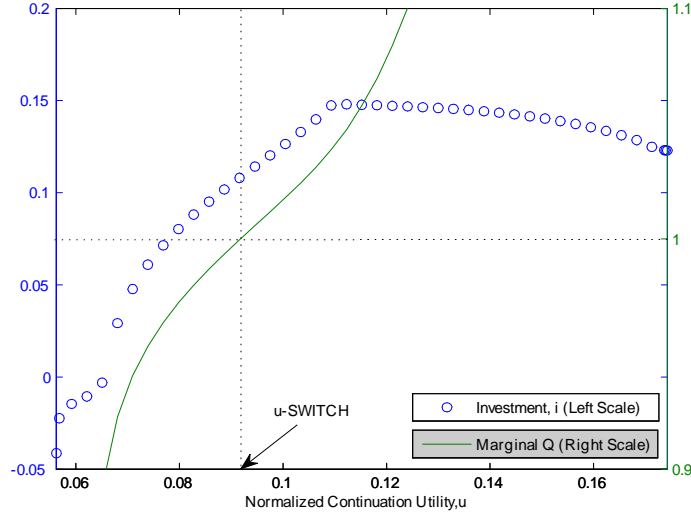


Figure 4: Marginal Q and Investment.

difference is that concavity of firms' value function arises in neoclassical model because of the concavity of the production function. Here the operating profit function is linear, and the concavity of the value function is induced by risk aversion of the manager.

The general pattern of investment and dividend policies is clear from Figure 5. Firms with high manager equity share invest more and pay fewer dividends. Due to higher investment rate, they must also grow faster on average. The next subsection discusses the dynamics of continuation utility, e.g. the manager's equity share under the optimal contract.

3.3 Dynamics of Continuation Utility

The evolution of continuation utility is the key to understand firm dynamics in the model. We first focus on the response of continuation utility with respect to observable shocks. We denote

$$\eta(K, u) \equiv -V_U(K, U).$$

Note that $V_U(K, U)$ is the partial derivative of firm's value with respect to the promised utility, U ; therefore $\eta(u, K)$ is the marginal cost of utility provision. Using the homogeneity property of the value function (19), we have $\eta(K, u) = -v'(u)u^\gamma K^\gamma$. We also use the notation

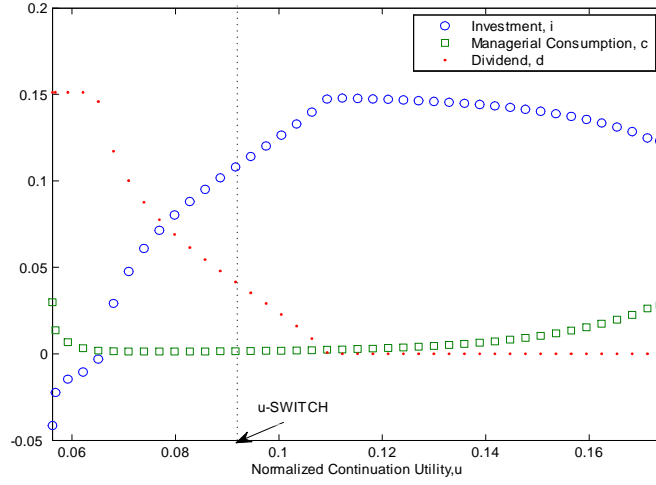


Figure 5: Managerial Consumption, Investment and Dividend Payout.

$Cov_t [d\eta(K_t, u_t), dB_{o,t}]$ to denote

$$\lim_{\Delta \rightarrow 0} \frac{1}{\Delta} Cov_t [\eta(K_{t+\Delta}, u_{t+\Delta}) - \eta(K_t, u_t), B_{o,t+\Delta} - B_{o,t}].$$

The following proposition states that under the optimal contract, the marginal cost of utility provision cannot respond to observable shocks.

Proposition 2 (Constancy of $\eta(u, K)$)

Under the optimal contract, the marginal cost of utility provision must not respond to observable shocks.⁷ Consequently,

$$Cov_t [d\eta(u_t, K_t), dB_{o,t}] = 0 \quad (25)$$

Proof. See Appendix A.6. ■

To understand the above proposition, note that the design of the continuation contract can be viewed as a process of assigning continuation utilities across time and states of nature. Optimality requires that the marginal cost of utility provision across states is equalized if continuation utility can be assigned in an unconstrained way. Although the principal must respect the incentive compatibility constraint when allocating continuation utilities across unobservable states, there is

⁷A similar condition is provided in Piskorski and Tchistyi (2011) in the case of risk-neutral agent, and in Li (2011) in the case of risk averse agent. Both papers consider the case of jump risk.

no constraint on doing so across observable states. In the absence of aggregate uncertainty, the relative price of consumption goods is constant across states. Since utility is additively separable, the marginal cost of utility across observable states must be constant. In continuous time, this is equivalent to the diffusion coefficient of $d\eta(u_t, K_t)$ on B_o being zero, as in (25).

Optimal compensation in our dynamic model differs from that in static models. Standard results from the static setting imply the use of relative performance evaluation in incentive compatible contracts (for example, [Holmström \(1982\)](#)), that is, the agents should not be rewarded for higher output due to observable exogenous shocks. This is no longer true in our dynamic setting. Optimality requires marginal cost of utility provision to be equalized across states, not continuation utility itself. In general, continuation utility U should not be equalized across observable states unless the value function $V_U(K, U)$ does not depend on K . In our case, the effect of observable shocks is persistent and affects the state variable K . As a result, it is optimal to adjust future continuation utility in response to these shocks even though they do not carry any information about the unobservable actions.⁸

Proposition 2 is also related to the "inverse Euler equation" of [Rogerson \(1985\)](#), [Spear and Srivastava \(1987\)](#) and [Goloso et al. \(2003\)](#). In fact, under the optimal contract, the marginal cost of utility provision must be constant across time as well. Formally, one can prove that the process $\{e^{(\beta+\kappa-r)t}V_U(K_t, U_t)\}_{t \in [\tau_0, \tau]}$ must be a martingale. In setting where consumption and effort are separable, $V_U(K_t, U_t)$ equals the inverse of the instantaneous marginal utility of consumption, and $\{e^{(\beta+\kappa-r)t}V_U(K_t, U_t)\}_{t \in [\tau_0, \tau]}$ being a martingale is nothing but a continuous time version of the "inverse Euler equation". In fact, Proposition 2 can be viewed as a generalization of the "inverse Euler equation". One can show that under the same separability conditions in [Goloso et al. \(2003\)](#), consumption of the agent must not respond to contemporaneous observable shocks.

Proposition 2 implies a simple characterization of the optimal policy g_o , which is summarized as follows.

Corollary 2 (Optimal Response to Observable Shocks) *The optimal choice of sensitivity of continuation utility with respect to observable shocks, g_o , is given by:*

$$g_o(u) = 1 - \frac{\gamma v'(u)}{\gamma v'(u) + uv''(u)}. \quad (26)$$

⁸Similar observations are also made by [Li \(2011\)](#) and [Hoffmann and Pfeil \(2010\)](#).

If the (un-normalized) value function $V(K, U)$ is strictly concave in K and U , then $g_o > 0$ for all $u \in [u_L, u_H]$.

In addition, $g_o(u) < 1$ if $\eta(K, u) > 0$ and $g_o(u) > 1$ if $\eta(K, u) < 0$.

Proof. See Appendix A.6. ■

Corollary 2 has several important implications. First, continuation utility responds positively to observable shocks. The reason for this result is that for a fixed U , the marginal cost of utility provision, $\eta(K, \frac{U}{K})$, is decreasing in K . Keeping U constant, an increase in K lowers manager's equity share u . Consequently, the marginal cost of utility provision is lower because it is easier to implement risk sharing plans. A positive observable shock increases K . To equalize the marginal cost of utility provision across states, it is optimal for the shareholder to assign a higher continuation utility in response to the positive observable shock. This means $g_o(u) > 0$. This feature of our model can be interpreted as optimal "pay for luck". Note that this is in contrast with the standard results obtained in the static setting, for example, [Holmström \(1982\)](#).

The second implication of the above proposition is that managers' equity share in the firm decreases with observable shocks whenever the marginal cost of utility provision is positive. Note that g_o is the sensitivity of U with respect to the observable shocks $dB_{o,t}$. In our formulation, the normalized utility u responding negatively to $dB_{o,t}$ is equivalent to $g_o(u) < 1$. This feature of our model inherits the properties of optimal risk sharing plans in the case without moral hazard. Without moral hazard, the optimal allocation is to provide the manager with a constant consumption stream. This arrangement means that the manager's consumption stream as a fraction of the firm's total cash flow must decrease after a positive shock. Corollary 2 implies that the above feature of the optimal risk sharing plan remains in the case with moral hazard.

To be more specific, normalized utility, u responding negatively to observable shocks is an implication of the risk aversion of the agent. Risk aversion means that, for a fixed u , the marginal cost of utility provision, $\eta(K, u)$ increases in K . Without risk aversion, both manager's utility function and the operating profit function of the firm are linear. As a result, the value function $V(K, U)$ is constant return to scale. This means that the marginal cost of provision does not depend on K as long as normalized utility u is kept constant. On the other hand, keeping u constant, risk aversion implies that the marginal utility of the manager diminishes when the firm size and therefore

manager's consumption increases. Or equivalently, the marginal cost of utility provision increases with K if u is kept constant. A positive observable shock increases K , $\eta(K, u)$ will increase if one raises U in a way to keep u constant. This experiment implies that after an observable shock raises K , U needs to increase, but not as much as K does. As a result, the normalized utility, u drops. In other words, if a firm's size increases due to an observable shock, optimal contract implies the manager's equity share should drop. If this mechanism is important, then we should observe manager's equity share to be higher in smaller firms. The calibrated version of our model does generate an inverse relation between firm size and managers' equity share for exactly this reason.

Thirdly, for exactly the same reason, managers' equity share increases with observable shocks if the marginal cost of utility provision is negative. As typical in dynamic contracting problems, the marginal cost of utility provision may be negative. This means that punishing the manager for bad performance may be so costly that it results in a lower value of the firm. In this case, after a positive observable shock raises K , the marginal benefit of utility provision, $-\eta(K, u)$ increases if u is kept constant. To equalize $\eta(K, u)$ across states, u needs to go up. That is, in this case, manager's share in the firm rises after a positive observable shock, $g_o(u) > 1$.

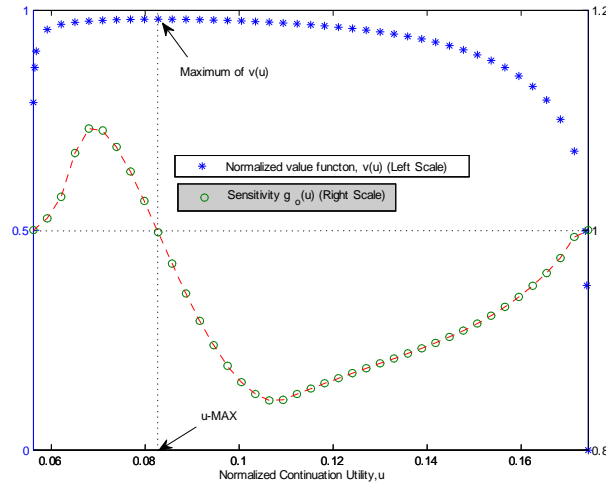


Figure 6: Optimal Sensitivity w.r.t Observable Shocks.

Figure 6 depicts the optimal sensitivity of continuation utility with respect to the observable shocks, g_o (circled line) and the normalized value function $v(u)$ (starred line) as functions of the normalized promised utility, u . All quantities are calculated under the parameter values we discuss

in the calibration section of the paper. u_{MAX} denote the maximizer of the value function. To the left of u_{MAX} , $v'(u) > 0$ and therefore $\eta(K, u) = -v'(u) u^\gamma K^\gamma < 0$. This is the region where to punish the agent, the principal has to suffer; therefore it benefits both parties to increase the promised utility of the agent. Keeping u constant, an increase in K is accompanied by a drop in the value of $\eta(K, u)$ as $-v'(u) u^\gamma < 0$. To equalize $\eta(K, u)$ across states, u must increase by concavity of $v(u)$. Here, $g_o(u) > 1$ and the normalized continuation utility rises after a positive observable shock.

In the case $u > u_{MAX}$, $\eta(K, u) = -v'(u) u^\gamma K^\gamma > 0$. Here the manager's equity share decreases with observable shocks. An increase in K increases $\eta(K, u)$ since $-v'(u) u^\gamma > 0$. To equalize $\eta(K, u)$ across states, u must drop because $v(u)$ is concave due to risk aversion.

The sensitivity of continuation utility with respect to unobservable shocks, $g_u(u)$ is shown in Figure 7, where g_u is marked with triangles. Note that $g_u(u) \geq g_o(u)$, that is, continuation utility is more sensitive to unobservable shocks, $dB_{u,t}$ than to observable shocks, $dB_{o,t}$. The intuition is clear, to provide incentives for the managers to invest, continuation utility must respond strongly to unobservable shocks. Just as in a standard principal-agent problems, there is a trade-off between risk sharing and incentive provision in our model. To provide incentive to the agent to invest, continuation utility needs to increase after a good unobservable shock, $dB_{u,t}$. More sensitivity implies higher incentives for the agent to invest; however, at the same time, it is also associated with higher welfare loss due to risk aversion and variations in the continuation utility. The optimal choice of g_u must trade off incentive provision against risk sharing with respect to unobservable shocks.

Finally, we note that over the interval (u_{MAX}, u_H) , a positive observable shock moves firms to the left and a positive unobservable shock moves firms to the right in the space of normalized utilities. If observable shocks are more important in accounting for the total volatility in capital accumulation, then we expect firm size as measured by K to be negatively correlated with promised utility, u . In this case risk sharing with respect to observable shocks is quantitatively more important than incentive provision, and good shocks are typically (when they are observable) associated a lower equity share of the manager in the firm. We show in Section 5 that this is indeed the case: In our model, small firms tend to invest more, pay fewer dividends, and grow faster than large firms.

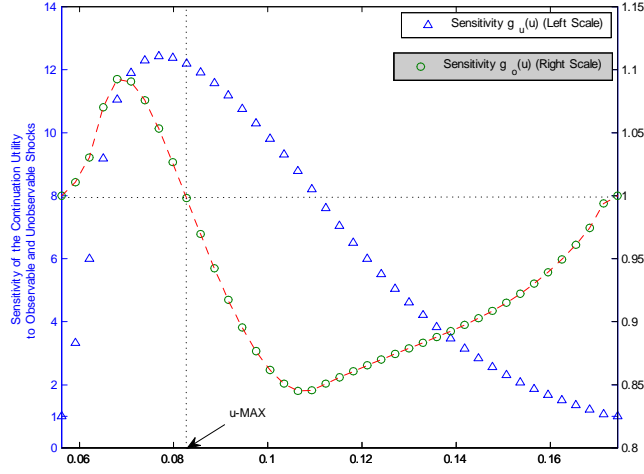


Figure 7: Optimal Sensitivity w.r.t Unobservable Shocks.

4 Aggregation and the Distribution of Firms

4.1 Aggregation and Numerical Procedure

In this section, we close the model by imposing the market clearing condition and verify our earlier conjecture on the existence of steady state where aggregate consumption of shareholders is constant. In our economy, firms' characteristics can be summarized by state variables (K, U) , or equivalently, (K, u) , where u is the normalized continuation utility of the manager as defined in (20). We use $\phi(K, u)$ to denote the stationary density of the joint distribution of firms at time t and use $\phi(u)$ to denote the density of the marginal distribution of u at time t . That is, $\phi(K, u)$ is the measure of firms with characteristic (K, u) in the steady-state distribution and $\phi(u)$ is the total measure of firms with normalized promised utility u in the steady-state distribution. Law of large numbers implies that these densities should integrate to $\frac{1}{\kappa}$ as firms enter the economy at rate 1 and dies at Poisson rate κ per unit of time.

The first result is of the initial condition on the size of new entrant firms.

Proposition 3 *The normalized initial promised utility to any manager is u_{SWITCH} and the total capital stock of all initial entrants at any time t is given by*

$$\int_{[U_L, U_H]} \left(\frac{U}{u_{SWITCH}} \right)^{\frac{1}{1-\gamma}} \Lambda(U) dU. \quad (27)$$

Note for each manager with reservation utility U , the manager choose the initial size of the tree. Optimality requires that the marginal benefit of investing in the tree to equal its marginal cost, which is 1 measured in current period consumption numeraire. That is, it is optimal to increase the size of a newly planted tree until $\frac{\partial}{\partial K} V(K, U) = 1$. This implies that the normalized promised utility of any manager upon entrance is u_{SWITCH} . Consequently, the initial capital stock given to a manager with reservation utility U is

$$K_0(U) = \left(\frac{U}{u_{SWITCH}} \right)^{\frac{1}{1-\gamma}}.$$

The total initial capital stock of all new entrant firms is therefore given by Equation (27) above.

We show in the appendix that the steady-state distribution exists. The stationary density $\phi(u)$ satisfies a version of the Kolmogorov forward equation, which we summarize in the following proposition.

Proposition 4 (Marginal Distribution of u) *On $(u_L, u_{SWITCH}) \cup (u_{SWITCH}, u_H)$, $\phi(u)$ satisfies*

$$\begin{aligned} 0 = & -\kappa\phi(u) - (1-\gamma)\frac{\partial}{\partial u} [\phi(u)u\mu_u(u)] \\ & + \frac{1}{2}(1-\gamma)^2 \frac{\partial^2}{\partial u^2} \left[\phi(u)u^2 \left[(g_u(u)-1)^2 \sigma_u^2 + (g_o(u)-1)^2 \sigma_o^2 \right] \right] \end{aligned}$$

on $(u_L, u_{SWITCH}) \cup (u_{SWITCH}, u_H)$.

The forward equation that describes the two-dimensional distribution $\phi(K, u)$ can be derived in a similar fashion. However, as noted by Ai (2011), in economies with homogeneous decision rules, the two dimensional distribution $\phi(K, u)$ can be summarized by a one dimensional measure $m(u)$ defined as follows:

$$m(u) = \int \phi(u, K) K dK. \quad (28)$$

The following proposition shows that $m(u)$ also satisfies a version of the Kolmogorov forward equation.

Proposition 5 (The Measure m) *Assume $\kappa + \delta - i(u) > 0$ for all $u \in [u_L, u_H]$, then $m(u)$ exists*

and satisfies

$$\begin{aligned}
0 = & -(\kappa + \delta - i(u)) m(u) \\
& - (1 - \gamma) \frac{\partial}{\partial u} \{m(u) u [\mu_u(u) + (g_u(u) - 1) \sigma_u^2 + (g_o(u) - 1) \sigma_o^2] dt\} \\
& + \frac{1}{2} (1 - \gamma)^2 \frac{\partial^2}{\partial u^2} \left[m(u) u^2 \left[(g_u(u) - 1)^2 \sigma_u^2 + (g_o(u) - 1)^2 \sigma_o^2 \right] \right].
\end{aligned}$$

on $(u_L, u_{SWITCH}) \cup (u_{SWITCH}, u_H)$.

Proofs of Proposition 4 and 5 can be found in Ai (2011). Using the definition of m , the total capital stock in the economy in steady-state is simply the integral of m :

$$\mathbf{K} = \int_{u_L}^{u_H} m(u) du. \quad (29)$$

Because allocation rules are homogeneous in our economy, aggregate output, aggregate consumption of managers, and aggregate investment in existing firms can all be written as integrals against the measure m :

$$\begin{aligned}
\int z K^\alpha N(K, U)^{1-\alpha} \Phi(K, U) dK dU &= \int_{u_L}^{u_H} z N(u)^{1-\alpha} m(u) du, \\
\int C(K, U) \Phi(K, U) dK dU &= \int_{u_L}^{u_H} c(u) m(u) du,
\end{aligned}$$

and

$$\int I(K, U) \Phi(K, U) dK dU = \int_{u_L}^{u_H} i(u) m(u) du,$$

respectively. Using the above, the market clearing conditions (16) and (17) can be written as:

$$\int_{u_L}^{u_H} n(u) m(u) du = 1,$$

and

$$\mathbf{C}^P + \int_{[u_L, u_H]} \left(\frac{U}{u} \right)^{\frac{1}{1-\gamma}} \Lambda(U) dU + \int_{u_L}^{u_H} [c(u) + i(u) + h(i(u))] m(u) du = \int_{u_L}^{u_H} z N(u)^{1-\alpha} m(u) du.$$

Note that the construction of measure m reduces the dimensionality of the cross-section distribution of firms and greatly simplifies the computation of equilibrium.

Using the above result, the competitive equilibrium can be calculated as follows.

1. Step 1: Starting from an initial guess of the marginal production of capital A , we solve the optimal contract and allocation rules by solving the ODE (24). Numerically, we use the Markov chain approximation method (Kushner and Dupuis (2001)) described in the appendix to solve the optimal control problem.
2. Step 2: After obtaining the policy function $c(u), i(u), g(u), h(u)$, we use Proposition 6 and 7 to construct the density $\phi(u)$ and the measure $m(u)$.
3. Step 3: We use measure m and Equation (29) to calculate the total capital stock in the steady state for the given A .
4. Step 4: We verify that A is the marginal product of capital using Equation (18). If $A > (<)$ $\alpha z \mathbf{K}^{\alpha-1}$, we choose a smaller (larger) A and resolve the contracts by repeating the above steps. We iterate this procedure until convergence, that is, until $A = \alpha z \mathbf{K}^{\alpha-1}$.

4.2 Distribution of Firms

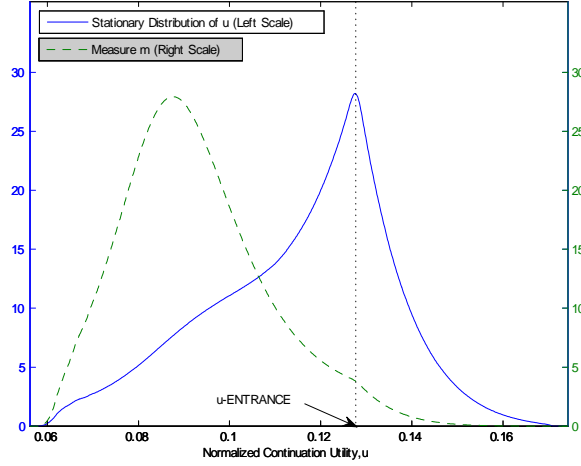


Figure 8: Density of Normalized Utility and the m-Measure.

Note that both of these measures have a mode at u_{SWITCH} (the label on the u -axis). As we remarked before, this is the point where the marginal product and the marginal cost of capital

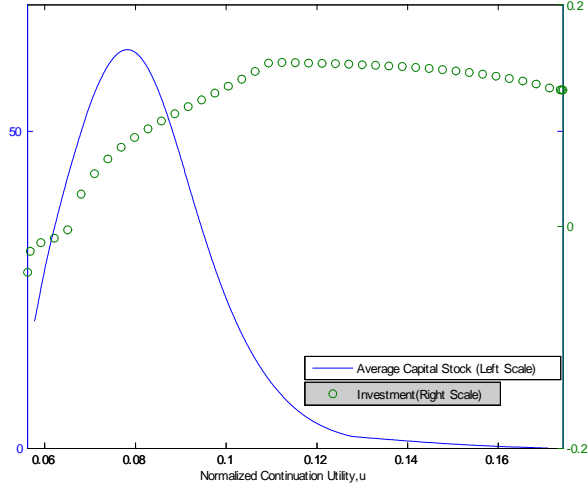


Figure 9: Average Size of Firms and the Investment Policy.

equalizes, and therefore firms have a tendency to converge to this point over time. In addition, Optimality of entrance implies that new firms enter into this economy at u_{SWITCH} as well. Note also, the majority of firms in this economy concentrated on the left of u_{MAX} . As explained earlier, points to the left of u_{MAX} is a “bad region” of the contract: an increase in u will simultaneously increase the utility of the manager, and the value of the firm. So firms tend to leave this region as soon as possible by choosing negative levels of investment.

By definition of $m(u)$ in Equation (27), the ratio $\frac{m(u)}{\phi(u)}$ is the average firm size at location u in steady state. In Figure 9, we plot the average size of firms (solid line, left scale) against normalized utility. We also plot the investment policy (circled line, right scale) on the same graph.

To the right of u_{MAX} , where most of the firms in the economy reside, average firm size decreases with investment. This pattern shows clearly our earlier claim that small firms invest more, grow faster, and pay fewer dividends.

5 Quantitative Implications of the Model

In this section, we calibrate our model to evaluate the quantitative implications of the model on the size distribution of the firms, and dependence of firm investment and payout policies on firm size. We choose the discount rate of the principal so that the interest rate in the economy is $r = 4\%$ in steady

state. We calibrate the death rate of firms to be $\kappa = 10\%$ per year, corresponding to the exit rate of firms reported in the Business Dynamics Statistics of US census. We choose the depreciate rate of firm specific capital $\delta = 1\%$, which amounts to assuming a 11% per depreciation rate of physical capital as in standard real business cycle models. We choose capital share $\alpha = 0.33$, which roughly matches the income share of capital and labor in US post war data. The productivity parameters z is calibrated to be 14.07, allowing our model to match a steady-state investment-output ratio in US postwar data of about 24%. We choose the risk aversion parameter of the manager to be $\gamma = 0.8$. Model implications are generally robust to the choice of the risk aversion parameter.

We choose the total volatility of firm-specific capital to be 36.0% per year, so that the volatility of output at the firm level roughly matches its empirical counterpart in the COMPUSTAT data. This number is lower than the volatility of sales reported in the US census data. Given that the tail of the size distribution is determined by large firms and large firms are substantially less volatile, it seems appropriate to be conservative and use COMPUSTAT data as a guidance here, because COMPUSTAT firms are publicly traded and are more representative of large firms in the US. We choose $\sigma_u = 0.8\%$ and $\sigma_o = 36\%$. As discussed above, it is important that σ_u is small relative to σ_o , so the risk sharing with respect to observable shocks is quantitatively important, and a favorable observable shock moves firms to the right in the normalized utility space. We simulate two millions firms from our model, and plot the statistics from the steady-state distribution of firms.

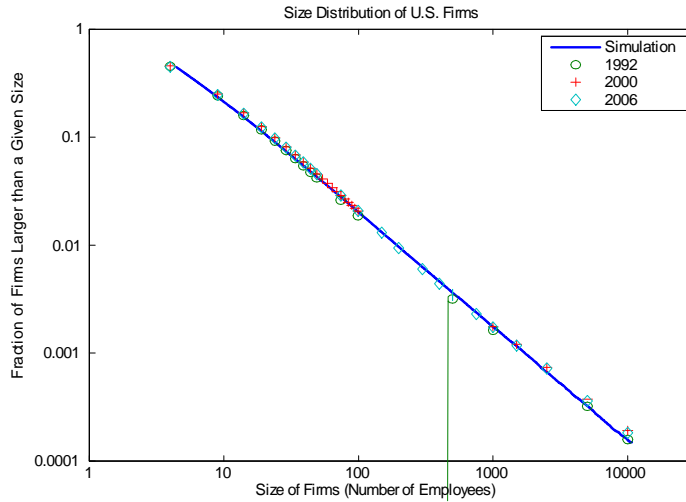


Figure 10: Size Distribution of Firms in the Data and Model.

Figure 10 plots, in log scale, the fraction of firms larger than a given size, K against $\log K$ in the data and in the model.⁹ Linearity of this curve is the defining characteristic of Pareto distributions. The size of firms is measured by number of employees as reported by Small Business Administration. We plot the distribution for all firms in 1992, 2000 and 2006, respectively. It is clear that the curve is very close to linear, and the slope is remarkably stable over time. The data set used to plot the figure contains the whole universe of firms reported in the Business Dynamics Statistics of US census. In the year 2006, for example, there are approximately 6 million firms. This remarkable pattern in the data has been documented by many previous researchers. The solid line in the figure is the same curve plotted from the steady-state of our calibrated model. Our model does a remarkable job in fitting the size distribution of firms in the data: it is almost indistinguishable from those plotted from data.

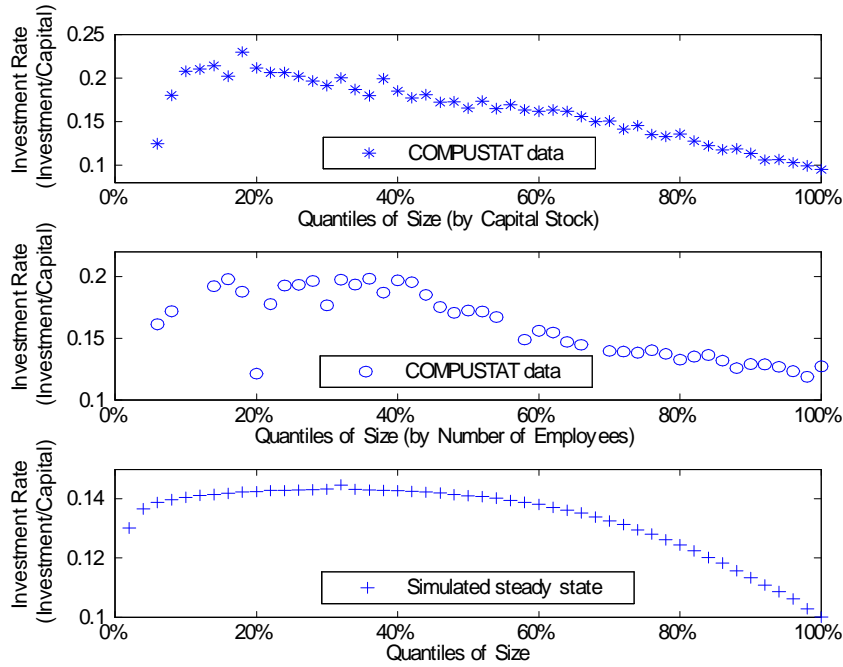


Figure 11: Investment to Capital Ratio Conditioning on Size.

We plot investment-to-capital ratio from the data and model in Figure 11. The top panel of Figure 11 is the plot of average investment rate against log firms size during the 1980-2006 period

⁹In the model, the size of a firm is determined up to a normalizing constant. We choose the normalization of K so that all lines have the same intercept.

in COMPUSTAT data. We measure the total capital stock of the firm as the total value of assets. In order to compare firm size across years, we adjust firm size by aggregate GDP from national product and income account published by U.S. Bureau of Economic Analysis. For each firm-year observations, we compute the investment rate (investment-to-capital ratio) and size of the firm as measured by total capital stock. We have a total of 300,000 observations. We divide firms into 50 size bins and plot the average investment rate of firms within each bin. Investment rate is clearly decreasing in firm size as shown in the figure. The middle panel of Figure 11 shows the investment rate-size relationship with size measured by total number of employees. Since comparing size of firms across years requires assumptions on the growth rate of the economy, we also experimented with the same figures using a single year data. The decreasing pattern of investment rate with size is robust across all specifications.

We plot in the bottom panel of Figure 11 the same graph using data generated by our model. The pattern of investment rate with respect to firm size resembles closely that in the data. The intuition for the decreasing pattern of investment rate with respect to size is as explained in the main text. Because $g_o(u) < 1$ in most of the region in the normalized utility, u -space, small firms are firms with high u . Marginal Q is higher in firms with high u , as the cost of utility provision is higher in these firms, and it is optimal to build up the capital stock so that the firm could move to the low u regions.

Why our model generates the fat tail of the size distribution of firms despite the low average investment rate of large firms? As shown by Luttmer (2011), in a model with constant expected growth rate, the tail index of the Pareto distribution is given by

$$\sqrt{\left(\frac{\mu}{\sigma^2}\right)^2 + \frac{2\kappa}{\sigma^2}} - \frac{\mu}{\sigma^2},$$

where μ is the expected growth rate and σ is the volatility of growth. The average investment rate of large firms is about 10% per year as show in figure 11. If all large firms have an identical investment rate of 10% per year, $\mu = i - \delta = 1\%$ per year. A back-of-envelope calculation shows that the tail index will be about 1.34, much higher than the 1.06 in the data. Although the average investment rate of large firms is low in our model, there are a substantial fraction of large firm invest at a high rate, close to 14% per year. In Figure 9, the average capital stock around the firm

entrance point u_{SWITCH} is fairly large despite the new entrance firms are small. This implies that there are a lot of large firms in this region as well. Firms are large in our economy for two reasons. Some of them have high investment rate, which are roughly in the region around u_{SWITCH} ; others grow large because they experienced a sequence of good shocks, which sends them to the left region in the utility space. Large firms with high investment rate are responsible for the fat tail of the size distribution of firms, and those with low investment rate are responsible for the observed low average investment conditioning on size. As a result, our model replicates the tail index of the empirical size distribution of firms despite that the investment behavior of firms is far complex than Gibrat's law.

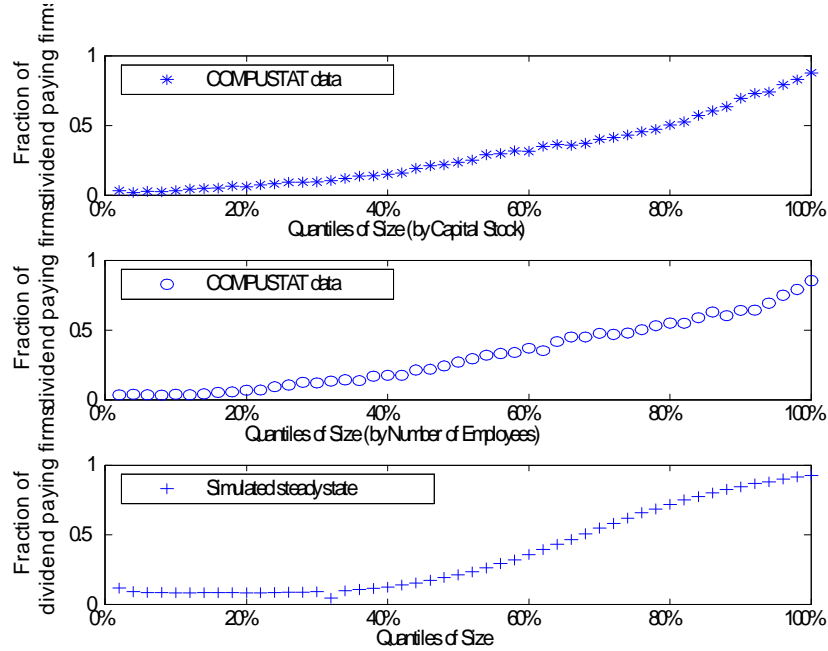


Figure 12: Fraction of Dividend Paying Firms Conditioning on Size.

Finally, in Figure 12, we plot the fraction of firms that pay dividend in each of the 50 size bins for COMPUSTAT firms 1980-2006, and the same from our model. It is clear that the fraction of firms that pay dividend is increasing in size, as measured by total capital stock (top panel), or by total number of employees (middle panel). Our model is largely consistent with the increasing pattern of probability of dividend payment in size as in the data.

6 Conclusion

We present a general equilibrium model with heterogeneous firms and moral hazard. Using continuous time methods, we solve for the optimal incentive contracts. Our model has predictions on the time series dynamics and the cross-sectional distribution of firms' investment, executive compensation and dividend payout policies. Theoretically, we provide a characterization of the optimal response of continuation utility with respect to observable shocks: the marginal cost of utility provision must not respond to observable shocks. Quantitatively, when calibrated to match standard macroeconomic moments, we show small firms invest more, pay fewer dividends and grow faster. More importantly, despite the heterogeneity in firms' decision rules, we show that the size distribution of firms in our model closely resembles a power law distribution with a slope coefficient close to 1.06, as in the data.

Small firms invest more, pay fewer dividends and grow faster. This phenomenon arises because the optimal contract optimally trades off incentive provision and risk sharing. Risk sharing implies manager's equity share must decrease with observable productivity shocks. As a result, manager's equity share is higher in small firms under the optimal contract. The presence of moral hazard problem implies that it is harder to provide incentives for managers to invest when they hold only a small equity share in the firm. Consequently, small firms invest more, pay fewer dividends and grow faster than large firms.

Power law requires large firms to grow fast enough to produce the fat tail of the size distribution. In our economy, firms become large for two reasons. Some of them grow large because of luck, i.e. because of a sequence of high productivity shocks. The managers' equity share in these firms decrease under the optimal contract. These firms contribute to the lower average investment of large firms. In steady state, a significant number of firms get large because of high historical investment, not productivity shocks. These firms continue to make high investment and grow fast. They contribute to the fat tail of the power law distribution.

References

- Dilip Abreu, David Pearce, and Ennio Stacchetti. Toward a theory of discounted repeated games with imperfect monitoring. *Econometrica*, 58:1041–1063, 1990.
- Hengjie Ai and Rui Li. Risk aversion, intertemporal elasticity of substitution and dynamic moral hazard. *Working paper, Duke University.*, 2011.
- Hengjie Ai. A note on aggregation in economies with homogenous decision rules. *Working paper, Duke University.*, 2011.
- Rui Albuquerque and Hugo Hopenhayn. Optimal lending contracts and firm dynamics. *Review of Economic Studies*, 71:285–315, 2004.
- Costas Arkolakis. A unified theory of firm selection and growth. *Working paper, Yale University*, 2011.
- Andrew Atkeson and Robert Lucas. On efficient distribution with private information. *Review of Economic Studies*, 59:427–453, 1992.
- Bruno Biais, Thomas Mariotti, and Jean-Charles Rochet Stéphane Villeneuve. Large risks, limited liability, and dynamic moral hazard. *Econometrica*, 78:73–118, 2010.
- Olivier Blanchard. Debt, deficits, and finite horizons. *Journal of Political Economy*, 93:223–247, 1985.
- Gian Clementi and Hugo Hopenhayn. A theory of financing constraints and firm dynamics. *Quarterly Journal of Economics*, 121:229–265, 2006.
- Gian Luca Clementi, Thomas Cooley, and Sonia Di Giannatale. A theory of firm decline. *Working paper, MIT*, 2008.
- Thomas Cooley and Vincenzo Quadrini. Financial markets and firm dynamics. *American Economic Review*, 91:1286–1310, 2001.
- Thomas Cooley, Ramon Marimon, and Vincenzo Quadrini. Aggregate consequences of limited contract enforceability. *Journal of Political Economy*, 112:817–847, 2004.

- Peter DeMarzo and Yuliy Sannikov. Optimal security design and dynamic capital structure in a continuous-time agency model. *Journal of Finance*, 61:2681–2724, 2006.
- Peter DeMarzo, Michael Fishman, Zhiguo He, and Neng Wang. Dynamic agency and q theory of investment. *Journal of Finance*, *Forthcoming*, 2009.
- David Evans. The relationship between firm growth, size, and age: Estimates for 100 manufacturing industries. *Journal of Industrial Economics*, 35:567–581, 1987.
- David Evans. Tests of alternative theories of firm growth. *Journal of Political Economy*, 95:657–674, 1987.
- Eugene Fama and Kenneth French. Disappearing dividends: changing firm characteristics or lower propensity to pay? *Journal of Financial Economics*, 60:3–43, 2001.
- Xavier Gabaix. Power laws in economics and finance. *Annual Review of Economics*, 1:255–293, 2009.
- Vito Gala and Brandon Julio. Convergence in corporate investments. *Working paper, London Business School*, 2011.
- Mikhail Golosov, Narayana Kocherlakota, and Aleh Tsyvinski. Optimal indirect and capital taxation. *Review of Economic Studies*, 70:569–587, 2003.
- Bronwyn Hall. The relationship between firm size and firm growth in the US manufacturing sector. *The journal of industrial economics*, 35:583–606, 1987.
- Fumio Hayashi. Tobin’s marginal q and average q: A neoclassical interpretation. *Econometrica*, 50:213–224, 1982.
- Zhiguo He. Optimal executive compensation when firm size follows geometric brownian motion. *Review of Financial Studies*, 22:859–892, 2009.
- Zhiguo He. A model of dynamic compensation and capital structure. *Journal of Financial Economics*, 100:351–366, 2011.
- Florian Hoffmann and Sebastian Pfeil. Reward for luck in a dynamic agency model. *Review of Financial Studies*, 23:3329–3345, 2010.

- Bengt Holmström. Moral hazard in teams. *Bell Journal of Economics*, 13:324–340, 1982.
- Hugo Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60:1127–1150, 1992.
- Boyan Jovanovic. Selection and the evolution of industry. *Econometrica*, 50:649–670, 1982.
- Jakob Klette and Samuel Kortum. Innovating firms and aggregate innovation. *Journal of Political Economy*, 112:986–1018, 2004.
- Harold Kushner and Paul Dupuis. *Numerical methods for stochastic control problems in continuous time*. Springer, 2001.
- Rui Li. Dynamic agency with persistent exogenous shocks. *Working paper, University of Wisconsin-Madison*, 2011.
- Erzo Luttmer. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122:1103–1144, 2007.
- Erzo Luttmer. Technology diffusion and growth. *Journal of Economic Theory*, *Forthcoming*, 2011.
- Edwin Mansfield. Entry, gibrat’s law, innovation, and the growth of firms. *American Economic Review*, 52:1023–1051, 1962.
- Roger Myerson. *Game Theory: Analysis of Conflict*. Harvard University Press, 1997.
- Tomasz Piskorski and Alexei Tchisty. Stochastic house appreciation and optimal mortgage lending. *Review of Financial Studies*, 24:1407–1446, 2011.
- Adriano Rampini and S Viswanathan. Collateral and capital structure. *Working paper, Duke University.*, 2012.
- William Rogerson. Repeated moral hazard. *Econometrica*, 53:69–76, 1985.
- Yuliy Sannikov. A continuous-time version of the principal-agent problem. *Review of Economic Studies*, 75:957–984, 2008.
- Stephen Spear and Sanjay Srivastava. On repeated moral hazard with discounting. *Review of Economic Studies*, 54:599–617, 1987.

Noah Williams. Persistent private information. *Econometrica*, 79:1233–1275, 2011.

$\{G_{ot}\}_{t \in [0, \tau]}$, such that

$$dU_t = \left(\beta \left(U_t - \frac{\tilde{C}_t^{1-\gamma}}{1-\gamma} \right) - \kappa (T(K_t) - U_t) \right) dt + G_{ut} \sigma_u dB_{u,t} + G_{ot} \sigma_o dB_{o,t} \text{ for } t \in [\tau_0, \tau]. \quad (33)$$

To characterize the random event of the arrival of the health shock of the manager, we define process $\{\varrho_t\}_{t \in [\tau_0, \tau]}$ as

$$\varrho_t = \begin{cases} 0 & \text{if } t < \tau \\ 1 & \text{if } t = \tau \end{cases}$$

and then $\{\varrho_t\}_{t \in [\tau_0, \tau]}$ is a simple jump process jumping from 0 to 1 with jump rate κ . It jumps when the health shock arrives. The associated compensated jump, $\{\Theta_t\}_{t \in [\tau_0, \tau]}$, martingale is

$$\Theta_t = -\kappa t + \varrho_t$$

and

$$d\Theta_t = -\kappa dt + \Delta\varrho_t \quad (34)$$

with $\Delta\varrho_t$ being the jump indicator. Therefore, all information is summarized by $\{B_{u,t}\}_{t \in [\tau_0, \tau]}$, $\{B_{o,t}\}_{t \in [\tau_0, \tau]}$ and $\{\Theta_t\}_{t \in [\tau_0, \tau]}$.

Now, for $t \leq \tau$, we define

$$\begin{aligned} \Gamma_t \left(\{C_s, I_s\}_{s \in [\tau_0, \tau]} \right) &= E_t \left[\beta \int_{\tau_0}^{\tau} e^{-\beta(s-\tau_0)} \frac{C_s^{1-\gamma}}{1-\gamma} ds + e^{-\beta\tau} T(K_\tau) | H^t \right] \\ &= \beta \int_{\tau_0}^t e^{-\beta(s-\tau_0)} \frac{C_s^{1-\gamma}}{1-\gamma} ds + e^{-\beta t} U_t. \end{aligned}$$

In words, $\{\Gamma_t\}_{t \in [\tau_0, \tau]}$ is the conditional expected total utility of the manager based on the information unfolded up to time t . Therefore, it is a H^t -adapted martingale. By martingale representation theorem, there exist three H^t -adapted square integrable processes, $\{G_{ut}\}_{t \in [\tau_0, \tau]}$, $\{G_{ot}\}_{t \in [\tau_0, \tau]}$ and $\{J_t\}_{t \in [\tau_0, \tau]}$, such that

$$d\Gamma_t = e^{-\beta t} G_{ut} \sigma_u dB_{u,t} + e^{-\beta t} G_{ot} \sigma_o dB_{o,t} + e^{-\beta t} J_t d\Theta_t \text{ for all } t \in [\tau_0, \tau].$$

On the other hand, according to the definition

$$d\Gamma_t = \beta e^{-\beta t} \frac{C_t^{1-\gamma}}{1-\gamma} dt + e^{-\beta t} dU_t - \beta e^{-\beta t} U_t dt \text{ for all } t \in [\tau_0, \tau].$$

Then we have

$$dU_t = \beta \left(U_t - \frac{C_t^{1-\gamma}}{1-\gamma} \right) dt + G_{ut} \sigma_u dB_{u,t} + G_{uot} \sigma_o dB_{o,t} + J_t d\Theta_t \text{ for all } t \in [\tau_0, \tau].$$

According to (34),

$$dU_t = \left(\beta \left(U_t - \frac{C_t^{1-\gamma}}{1-\gamma} \right) - \kappa J_t \right) dt + G_{ut} \sigma_u dB_{u,t} + G_{uot} \sigma_o dB_{o,t} + J_t \Delta \varrho_t \text{ for all } t \in [\tau_0, \tau].$$

If the health shock hits at t , the jump distance of the promised utility of the manager is $T(K_t) - U_t$.

Therefore $J_t = T(K_t) - U_t$ Then we have (33).

A.3 Proof of Lemma 1

For convenience, given capital \hat{K} , we define

$$\mathcal{M}(\hat{D}, \hat{K}) = \left\{ (\hat{C}, \hat{I}) \in \mathbb{R}_+ \times [I_L, \tau] : \hat{C} + \hat{I} + H(\hat{I}, \hat{K}) = \hat{D} \right\}.$$

Note that, according to the definition, an allocation rule (C, I, D, G_u, G_o) is feasible if and only if $(C_t, I_t) \in \mathcal{M}(D_t)$ for all $t \in [\tau_0, \tau]$. According to (5), $\mathcal{M}(\hat{D})$ can be rewritten as

$$\mathcal{M}(\hat{D}, \hat{K}) = \left\{ (\hat{C}, \hat{I}) \in \mathbb{R}^2 : \hat{I} \in [I_1(\hat{D}, \hat{K}), I_2(\hat{D}, \hat{K})] \text{ and } \hat{C} = \hat{D} - \hat{I} - H(\hat{I}, \hat{K}) \right\}$$

with

$$I_1(\hat{D}, \hat{K}) = \max \left\{ I_L, \frac{\omega i^* - 1 - \sqrt{1 - 2\omega i^* + 4\omega \frac{\hat{D}}{\hat{K}}}}{2\omega} \right\}$$

$$I_2(\hat{D}, \hat{K}) = \frac{\omega i^* - 1 + \sqrt{1 - 2\omega i^* + 4\omega \frac{\hat{D}}{\hat{K}}}}{2\omega}.$$

The investment and compensation plan is *interior* if $I_t \in (I_1(D_t, K_t), I_2(D_t, K_t))$ for all $t \in [\tau_0, \tau]$.

We utilize the method applied in [Sannikov \(2008\)](#) to prove (14).

Suppose not, such that (C, I) does not satisfy condition (14) over time horizon $[\tau_0, t]$ with $t \in [\tau_0, \tau)$ and strictly positive probability. Let $\{U_t\}_{t \in [\tau_0, \tau)}$ be the continuation utility process generated by (C, I) under the allocation rule. We define

$$\hat{\phi}_t(C', I') = \beta \int_{\tau_0}^t e^{-\beta(s-\tau_0)} \frac{C_s'^{1-\gamma}}{1-\gamma} ds + e^{-\beta t} U_t \text{ for } t \in [\tau_0, \tau)$$

where (C', I') is some alternative investment-consumption plan. So, $\hat{\phi}_t(C', I')$ is the conditional expected utility of the manager based on history H^t if he adopts (C', I') from τ_0 to t and then switches to (C, I) . Obviously, $\hat{\phi}_{\tau_0}(C', I') = U_0$, the expected utility of the manager if he chooses (C, I) from the beginning. According to equation (9), the motion of the continuation utility, we have

$$\begin{aligned} d\hat{\phi}_t(C', I') &= \beta e^{-\beta t} \frac{C_t'^{1-\gamma}}{1-\gamma} dt - \beta e^{-\beta t} U_t dt + e^{-\beta t} dU_t \\ &= e^{-\beta t} \beta \left[\frac{C_t'^{1-\gamma}}{1-\gamma} - \frac{C_t^{1-\gamma}}{1-\gamma} \right] dt + e^{-\beta t} \frac{G_{ut}}{K_t} K_t \sigma_u dB_{u,t} + e^{-\beta t} \frac{G_{ot}}{K_t} K_t \sigma_o dB_{o,t} + e^{-\beta t} J_t d\Theta_t. \end{aligned}$$

Recall that $\{B_{u,t}\}_{t \in [\tau_0, \tau]}$ is a standard Brownian motion under the measure \mathcal{G} generated by (C, I) . Similarly, let $\{B'_{u,t}\}_{t \in [\tau_0, \tau]}$ be a standard Brownian motion under the measure \mathcal{G}' , which is generated by the alternative plan (C', I') . The transfer between the two measure is according to *Girsanov theorem*. Now we have

$$dB_{u,t} = dB'_{u,t} + \frac{1}{K_t \sigma_u} (I'_t - I_t) dt$$

and then

$$d\hat{\phi}_t(C', I') = e^{-\beta t} \left\{ \left[\beta \frac{C_t'^{1-\gamma}}{1-\gamma} - \beta \frac{C_t^{1-\gamma}}{1-\gamma} + \frac{G_{ut}}{K_t} (I'_t - I_t) \right] dt + \frac{G_{ut}}{K_t} K_t \sigma_u dB'_{u,t} + \frac{G_{ot}}{K_t} K_t \sigma_o dB_{o,t} + e^{-\beta t} J_t d\Theta_t \right\}.$$

Now, we select (C', I') such that

$$(C'_t, I'_t) \in \arg \max_{(\tilde{C}, \tilde{I}) \in \mathcal{M}(D_t, K_t)} \beta \frac{\tilde{C}^{1-\gamma}}{1-\gamma} - \frac{G_{ut}}{K_t} I_t \text{ for } t \in [\tau_0, \tau).$$

Then the drift of $\left\{\hat{\phi}_t(C', I')\right\}_{t \in [\tau_0, \tau]}$ is strictly positive and it is a sub-martingale under measure \mathcal{G}' . Therefore

$$E' \left[\hat{\phi}_{\hat{t}}(C', I') \right] > \hat{\phi}_{\tau_0}(C', I') = U_0.$$

Here, $E'[\cdot]$ is the expectation based on \mathcal{G}' . So, (C, I) is dominated by the plan that adopting (C', I') from the begging and switching to (C, I) at \hat{t} . We have a contradiction to the fact that (C, I) is optimal.

On the other hand, suppose (C, I) satisfies (14), then $\left\{\hat{\phi}_t(C', I')\right\}_{t \in [\tau_0, \tau]}$ is a super-martingale under \mathcal{G}' for any alternative plan (C', I') . Denote the overall utility generated by plan (C', I')

$$\hat{\phi}_\tau(C', I') = \beta \int_{\tau_0}^{\tau} e^{-\beta(t-\tau_0)} \beta \frac{C_t'^{1-\gamma}}{1-\gamma} dt.$$

Note that, given any feasible plan, the expected continuation utility of the manager is bounded by $[U_L, U_H]$. Therefore

$$U_0 = \hat{\phi}_0(C', I') \geq E' \left[\hat{\phi}_\tau(C', I') \right]$$

and (C, I) weakly dominates (C', I') .

Now, we suppose that (C, I) is interior. Since $C_t = D_t - I_t - H(I_t, K_t)$, and then according to (14), I_t solves the problem

$$\max_{\tilde{I}} \beta \frac{\left(D_t - \tilde{I} - H(\tilde{I}, K_t) \right)^{1-\gamma}}{1-\gamma} - \frac{G_{ut}}{K_t} \tilde{I}$$

and it is easy to check that the objective function is concave. So by taking the first order condition we have the desired result.

A.4 The Lower Bound of the Normalized Value Function

In this section, we compute a lower bound of the normalized value function for three purposes: First, this helps us understand the basic trade off between incentive provision and risk sharing we focus on in this paper; second, it helps us to compute the left and right boundary points of the normalized value function; third, it will also help us to motivate our assumptions on the parameters values of the model.

The lower bound on the principal's value function can be constructed by identifying a simple form of incentive compatible plan. Consider the following class of compensation contracts. The total payment from the principal to the agent is ξAK_t , where $0 \leq \xi \leq 1$, and the rest of output is paid out to shareholders as dividend. Given ξ , the agent chooses optimally the division of ξAK_t between his own consumption and investment in the firm. We call a plan of this form a *constant-share* plan. As we vary ξ from 0 to 1, this procedure traces out the market value of a firm as a function of the utility delivered to the manager. We denote this value function as $v_{LB}(u)$, where u is the normalized utility as discussed in the above section. Note $v_{LB}(u)$ provides a lower bound to the value function of the principal, as a constant-share plans are incentive compatible by construction, but not necessarily optimal.

It is convenient for us to impose a restriction on terminal utility parameter u_B . We assume $u_B \in [\bar{u}_B, \underline{u}_B]$ with \bar{u}_B being the normalized expected utility of the manager if he owns the full share of the tree and the tree is never hit by the disease, namely, $\kappa = 0$; \underline{u}_B being that if the manager owns zero share of the tree and the tree is never hit by the disease.

It is convenient to introduce the following notation.

$$\hat{\beta} = \beta + (1 - \gamma) \delta + \frac{1}{2} \gamma (1 - \gamma) [\sigma_u^2 + \sigma_o^2]. \quad (35)$$

We first make the following assumption on $\hat{\beta}$. If the adjustment cost is 0 the following assumption guarantees that the lower and upper bound, \underline{u}_B and \bar{u}_B , of u_B exists.

Assumption 1

$$\hat{\beta} = \beta + (1 - \gamma) \delta + \frac{1}{2} \gamma (1 - \gamma) [\sigma_u^2 + \sigma_o^2] > 0 \quad (36)$$

In Appendix A.4.1, we provide a condition under which the \underline{u}_B and \bar{u}_B exists if the adjustment cost is not zero. We also show that if the lower and upper bound of u_B exists, the lower bound of the principal's value function, $v_{LB}(u)$, exists and uniquely characterized by a HJB equation. First we discuss the existence of the \bar{u}_B and \underline{u}_B .

A.4.1 The Upper and Lower Bounds \bar{u}_B and \underline{u}_B

Let $p_d \in \{0, 1\}$ and $\kappa = 0$. Let $U^B(K, 1)$ and $U^B(K, 0)$ be the un-normalized expected utility associated with the contracts that generate \bar{u}_B and \underline{u}_B according to their definition. Here, 1 and

0 stand for the manager's share. So we define $p_d \in \{0, 1\}$. Then $U^B(K, p_d)$ satisfies the following HJB equation:

$$\begin{aligned} 0 = & \max_i \beta \left(\frac{((p_d A - \varphi(i)) K)^{1-\gamma}}{1-\gamma} - U^B(K, p_d) \right) + K(i - \delta) U_K^B(K, p_d) \\ & + \frac{1}{2} K^2 (\sigma_u^2 + \sigma_o^2) U_{KK}^B(K, p_d). \end{aligned}$$

KK

satisfies the following HJB equation:

$$\begin{aligned} 0 = & \max_i \beta \left(\frac{((p_d A - \varphi(i)) K)^{1-\gamma}}{1-\gamma} - U^{LB}(K, p_d) \right) + K(i - \delta) U_K^{LB}(K, p_d) \\ & + \frac{1}{2} K^2 (\sigma_u^2 + \sigma_o^2) U_{KK}^{LB}(K, p_d) + \kappa \left(\frac{(u_T K)^{1-\gamma}}{1-\gamma} - U^{LB}(K, p_d) \right). \end{aligned}$$

Similar with what we did, we guess

$$U^{LB}(K, p_d) = \frac{(u^{LB}(p_d) K)^{1-\gamma}}{1-\gamma}.$$

And then we have for $p_d \in [0, 1]$, $u^{LB}(p_d)$ satisfies the following HJB equation.

$$0 = \max_i \frac{1}{1-\gamma} \left[\beta \frac{(p_d A - \varphi(i))^{1-\gamma}}{(u^{LB}(p_d))^{1-\gamma}} + (1-\gamma) i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(p_d)} \right)^{1-\gamma} - 1 \right] \right] \quad (40)$$

First, we show that for $p_d = 0$ and 1 the equation has a unique solution in $[\underline{u}_B, \bar{u}_B]$.

If $p_d = 0$, (40) is rewritten as

$$0 = \max_i \frac{1}{1-\gamma} \left[\beta \frac{(-\varphi(i))^{1-\gamma}}{(u^{LB}(p_d))^{1-\gamma}} + (1-\gamma) i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(p_d)} \right)^{1-\gamma} - 1 \right] \right]$$

Note that the right hand side of the equation, which is denoted as $\chi^0(u^{LB}(p_d))$, is continuous in $u^{LB}(p_d)$ over $[\underline{u}_B, \bar{u}_B]$. Since we know

$$\begin{aligned} \chi^0(\underline{u}_B) &= \max_i \frac{1}{1-\gamma} \left[\beta \frac{(-\varphi(i))^{1-\gamma}}{(\underline{u}_B)^{1-\gamma}} + (1-\gamma) i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{\underline{u}_B} \right)^{1-\gamma} - 1 \right] \right] \\ &\geq \max_i \frac{1}{1-\gamma} \left[\beta \frac{(-\varphi(i))^{1-\gamma}}{(\underline{u}_B)^{1-\gamma}} + (1-\gamma) i - \hat{\beta} \right] = 0. \end{aligned}$$

The inequality above is due to the assumption that $u_T \in [\underline{u}_B, \bar{u}_B]$ and the last equation is due to the HJB equation that \underline{u}_B satisfies. On the other hand

$$\begin{aligned} \chi^0(\bar{u}_B) &= \max_i \frac{1}{1-\gamma} \left[\beta \frac{(-\varphi(i))^{1-\gamma}}{(\bar{u}_B)^{1-\gamma}} + (1-\gamma) i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{\bar{u}_B} \right)^{1-\gamma} - 1 \right] \right] \\ &\leq \max_i \frac{1}{1-\gamma} \left[\beta \frac{(A - \varphi(i))^{1-\gamma}}{(\bar{u}_B)^{1-\gamma}} + (1-\gamma) i - \hat{\beta} \right] = 0. \end{aligned}$$

The inequality above is due to the fact that $A > 0$ and $u_T \in [\underline{u}_B, \bar{u}_B]$ and the last equation is due to the HJB equation that \bar{u}_B satisfies. Since $\chi^0(\cdot)$ is continuous on $[\underline{u}_B, \bar{u}_B]$. Then, we choose the largest solution $u^{LB}(0) \in [\underline{u}_B, \bar{u}_B]$ such that $\chi^0(u^{LB}(0)) = 0$. Similarly, we can prove there is a unique maximal solution $u^{LB}(1) \in [\underline{u}_B, \bar{u}_B]$.

Now, for any $p_d \in (0, 1)$, we denote the right hand side of (40) as $\chi^{p_d}(u^{LB}(p_d))$. Note that

$$\begin{aligned}\chi^{p_d}(u^{LB}(0)) &= \max_i \frac{1}{1-\gamma} \left[\beta \frac{(p_d A - \varphi(i))^{1-\gamma}}{(u^{LB}(0))^{1-\gamma}} + (1-\gamma)i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(0)} \right)^{1-\gamma} - 1 \right] \right] \\ &\geq \max_i \frac{1}{1-\gamma} \left[\beta \frac{(-\varphi(i))^{1-\gamma}}{(u^{LB}(0))^{1-\gamma}} + (1-\gamma)i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(0)} \right)^{1-\gamma} - 1 \right] \right] = 0.\end{aligned}$$

The inequality above is due to the fact that $p_d > 0$ and the equality is due to the HJB equation that $u^{LB}(0)$ satisfies. On the other hand

$$\begin{aligned}\chi^{p_d}(u^{LB}(1)) &= \max_i \frac{1}{1-\gamma} \left[\beta \frac{(p_d A - \varphi(i))^{1-\gamma}}{(u^{LB}(1))^{1-\gamma}} + (1-\gamma)i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(1)} \right)^{1-\gamma} - 1 \right] \right] \\ &\leq \max_i \frac{1}{1-\gamma} \left[\beta \frac{(A - \varphi(i))^{1-\gamma}}{(u^{LB}(1))^{1-\gamma}} + (1-\gamma)i - \hat{\beta} + \kappa \left[\left(\frac{u_T}{u^{LB}(1)} \right)^{1-\gamma} - 1 \right] \right] = 0\end{aligned}$$

The inequality above is due to the fact that $p_d < 1$ and the equality is due to the HJB equation that $u^{LB}(1)$ satisfies. Since $\chi^{p_d}(\cdot)$ is continuous over $[u^{LB}(0), u^{LB}(1)]$. Then we have an unique maximal solution $u^{LB}(p_d)$ which satisfies (40).

A.5 Proof of Proposition 1

We only need to derive the HJB equation.

First, we derive the HJB equation that the value function $V(K, U)$ satisfies. For $t \in [0, \tau)$, according to the law of motion of the state variables, (33) and (10), by Ito's lemma we have $V(K, U)$ satisfies the following differential equation

$$\begin{aligned}\max_{C, I, D, G_u, G_o} & D - (\beta + \kappa) V(K, U) + V_K(K, U) (I - \delta K) + V_U(K, U) \left(\beta \left(U - \frac{C^{1-\gamma}}{1-\gamma} \right) - \kappa (T(K) - U) \right) \\ & + \frac{1}{2} \left[V_{KK}(K, U) K^2 (\sigma_u^2 + \sigma_o^2) + V_{UU}(K, U) (G_u^2 \sigma_u^2 + G_o^2 \sigma_o^2) \right. \\ & \quad \left. + 2V_{KU}(K, U) (G_u \sigma_u^2 + G_o \sigma_o^2) \right].\end{aligned}\tag{41}$$

subject to

$$0 \leq C(K, U) + I(K, U) + H(I, K) + D(K, U) \leq AK$$

and (14). According to the normalization defined by (19) and (20), we have

$$V_K(K, U) = v(u) - uv'(u); \quad (42)$$

$$V_U(K, U) = \frac{1}{1-\gamma} \frac{K}{U} uv'(u);$$

$$V_{KK}(K, U) = \frac{1}{K} u^2 v''(u); \quad (43)$$

$$V_{UU}(K, U) = \frac{1}{(1-\gamma)^2} \frac{K}{U^2} [\gamma uv'(u) + u^2 v''(u)]; \quad (44)$$

$$V_{KU}(K, U) = -u^2 v''(u) \frac{1}{1-\gamma} \frac{1}{U}. \quad (45)$$

By plugging in the expressions above into (41), we have (24).

A.6 Proof of Proposition 2

The expression of h , equation (26) is directly derived from the HJB equation (23), according to which h solves the following problem

$$\max_{\hat{h}} \frac{1}{2} uv'(u) (\gamma \hat{h}^2 \sigma_o^2) + \frac{1}{2} u^2 v''(u) \left[(\hat{h} - 1)^2 \sigma_o^2 \right].$$

the first order condition implies

$$h(u) = \frac{uv''(u)}{\gamma v'(u) + uv''(u)}.$$

Then we have the expression of $h(u)$. By (43), we have that $v''(u)$ has the same sign as $V_{KK}(K, U)$ which is strictly negative. By (44), $\gamma v'(u) + uv''(u)$ has the same sign as $V_{UU}(K, U)$. Then we have the result. On the other hand, the sensitivity of the normalized continuation utility to observable shocks is

$$h(u) - 1 = \frac{\gamma v'(u)}{\gamma v'(u) + uv''(u)}.$$

Since $v'(u)$ is increasing in u . And $h(u) - 1 = 0$ if and only if $v'(u) = 0$, say $v(\cdot)$ reaches its maximum. Then we have the desired result.

Now we prove Proposition, By Ito's formula we have the law of motion of $V_U(U, K)$ as follows

¹²,

$$\begin{aligned}
dV_U(U, K) = & V_{UU}(U, K) \left[\beta \left(U - \frac{C^{1-\gamma}}{1-\gamma} \right) - \kappa (T(K) - U) dt + G_{ut} \sigma_u dB_{u,t} + G_{ot} \sigma_o dB_{o,t} \right] \\
& + V_{UK}(U, K) [(-\delta + i) K dt + K \sigma_u dB_{u,t} + K \sigma_o dB_{o,t}] \\
& + \frac{1}{2} [V_{UUU} (G_{ut}^2 \sigma_u^2 + G_{ot}^2 \sigma_o^2) dt + V_{UKK} K^2 (\sigma_u^2 + \sigma_o^2) + 2V_{UKU} (G_{ut} \sigma_u^2 + G_{ot} \sigma_o^2)] \quad (46)
\end{aligned}$$

The coefficient of $dB_{o,t}$ on the right hand side is $V_{UU}(U, K) G_{ot} \sigma_o + V_{UK}(U, K) K \sigma_o$. By (44), (44), the expression of $h(u)$ and the way we normalize the sensitivity, we have

$$\begin{aligned}
V_{UU}(U, K) G_{ot} \sigma_o + V_{UK}(U, K) K \sigma_o &= \frac{K}{(1-\gamma)U} [u^2 v''(u) - (\gamma u v'(u) + u^2 v''(u)) h(u)] \\
&= 0.
\end{aligned}$$

Therefore, we have the first result.

From the right hand side of (46), we can see that the drift of $V_U(U, K)$ is

$$\begin{aligned}
V_{UU}(U, K) &= \beta \left(U - \frac{C^{1-\gamma}}{1-\gamma} \right) - \kappa (T(K) - U) + V_{UK}(U, K) (-\delta + i) K \\
&+ \frac{1}{2} [V_{UUU} (G_{ut}^2 \sigma_u^2 + G_{ot}^2 \sigma_o^2) dt + V_{UKK} K^2 (\sigma_u^2 + \sigma_o^2) + 2V_{UKU} (G_{ut} \sigma_u^2 + G_{ot} \sigma_o^2)].
\end{aligned}$$

By envelope theorem, we take the first order derivative with respect to U on both hand sides of HJB equation (41) and denote the optimal polices by C^* , I^* , D^* , G_u^* and G_o^* we have

$$\begin{aligned}
0 &= D^* - (\beta + \kappa) V_U(K, U) + V_{UK}(K, U) (I^* - \delta K) + V_{UU}(K, U) \left(\beta \left(U - \frac{C^{*1-\gamma}}{1-\gamma} \right) - \kappa (T(K) - U) \right) \\
&+ (\beta + \kappa) V_U(K, U) \\
&+ \frac{1}{2} [V_{UUU} (G_{ut}^{*2} \sigma_u^2 + G_{ot}^{*2} \sigma_o^2) dt + V_{UKK} K^2 (\sigma_u^2 + \sigma_o^2) + 2V_{UKU} (G_{ut}^* \sigma_u^2 + G_{ot}^* \sigma_o^2)] \\
&= V_{UU}(U, K) \beta \left(U - \frac{C^{*1-\gamma}}{1-\gamma} \right) - \kappa (T(K) - U) + V_{UK}(U, K) (I^* - \delta K) \\
&+ \frac{1}{2} [V_{UUU} (G_{ut}^{*2} \sigma_u^2 + G_{ot}^{*2} \sigma_o^2) dt + V_{UKK} K^2 (\sigma_u^2 + \sigma_o^2) + 2V_{UKU} (G_{ut}^* \sigma_u^2 + G_{ot}^* \sigma_o^2)].
\end{aligned}$$

Then we have the drift of $V_U(U, K)$ is 0.

¹²For convience, we omitte the arguments K, U in the policy functions.

A.7 Proof of Proposition 1

According to the HJB differential equation (23), if there is no adjustment cost, c and i jointly solve the following problem

$$\max_{c,i} \left(v(u) - uv'(u) - 1 \right) i - uv'(u) \frac{\beta \left(\frac{c}{u} \right)^{1-\gamma}}{1-\gamma}.$$

By Lemma 3, if $u \in (u_{SWITCH}, u_H)$, $v(u) - uv'(u) - 1 > 0$. Then $c + i \leq A$ must be binding. Then the dividend paid to the shareholders is 0; if $u \in (u_L, SWITCH)$, $v(u) - uv'(u) - 1 < 0$. Then $c + i \geq 0$ must be binding. Then the dividend paid to the shareholders is A .

Suppose, in addition, we assume that $v(u) \leq 1$ for all u . We have $v'(u_{SWITCH}) < 0$. Note that $v'(u_{MAX}) = 0$. Due to the concavity of $v(\cdot)$, which is implied by the concavity of $V(\cdot, \cdot)$, we have

$$u_{SWITCH} > u_{MAX}.$$

In the other direction, if $u_{SWITCH} > u_{MAX}$ holds, then Lemma 3 implies that $v(u_{MAX}) - u_{MAX}v'(u_{MAX}) < 1$ and then we have the result.