Spring 5-15-2020

Essays on Macro and Financial Economics

Linyi Cao
Washington University in St. Louis

WASHINGTON UNIVERSITY IN ST. LOUIS

Department of Economics

Dissertation Examination Committee:
Rodolfo Manuelli, Chair
Gaetano Antinolfi
Francisco Buera
Yongseok Shin
Guillaume Vandenbroucke

Essays on Macro and Financial Economics
by
Linyi Cao

A dissertation presented to
The Graduate School
of Washington University in
partial fulfillment of the
requirements for the degree
of Doctor of Philosophy

May, 2020
St. Louis, Missouri

Table of Contents

List of Figures

List of Tables

Acknowledgements

I am deeply indebted to Rodolfo Manuelli for his patience and continuous support during my years as a PhD student. His guidance and insights have undoubtedly influenced the present dissertation. I also thank Yongseok Shin, Gaetano Antinolfi for various discussions on the main chapters of the dissertation. Those chapters were further enriched by comments and suggestions by Francisco Buera, Guillaume Vandenbroucke, Michele Boldrin, and Cheng Wang whom I thank for their involvement too. Finally, I thank the Graduate School of Arts and Sciences of Washington University for the financial support.

Linyi Cao

Washington University in St. Louis

May, 2020

ABSTRACT OF THE DISSERTATION

Essays on Macro and Financial Economics

by

Linyi Cao

Doctor of Philosophy in Economics

Washington University in St. Louis, 2020

Professor Rodolfo Manuelli, Chair

Nurturing Young Public Firms over Real Business Cycles. In this paper, I develop a theory of financial intermediation in a general equilibrium environment, to study the interactions between households, financial intermediation, and entrepreneurs over real business cycles. In the model, the financial intermediary, who resembles real-life private equity (PE) groups and investment bankers, works as a nurturer of young public firms. It performs screening and sorting on entrepreneurs, then allocates resources to them, borrowed from the households. However, the effort intensity in screening decreases when the financial intermediary is flooded by resources, so do the average quality of financial services and the commission rate, which predict the countercyclicality of those variables. This countercyclicality of efficiency in the financial sector promises a dampening effect on economic volatility. I use the U.S. initial public offering (IPO) data, as well as selected PE data to document that the commission rate is indeed countercyclical. Its correlation with the cyclical component of total output is around -0.21. I calibrate the model to the U.S. financial market and conduct several counterfactual exercises. I find that a 20% drop in the financial intermediary's cost of effort dampens the total output volatility by 0.24% and the household consumption volatility by 0.53%. While a binding commission rate cap amplifies the volatility by 0.36% and 0.54% respectively.

Antitrust Policy in a Globalized Economy. Antitrust policies have been relaxed and the number of mergers and acquisitions (M&A) has risen rapidly since the 1980s in the United States. This paper provides a framework to evaluate the cost and benefits of antitrust policy in a global context. M&A reallocate resources from small to large and typically more productive firms, while also increasing their monopoly power. An optimal antitrust policy seeks a balance between the positive productivity effect and the negative markup effect. In a globalized economy, increasing productivity fully accrues to domestic firms/workers while a higher markup only partially hurts domestic consumers. The weakening antitrust policy since the 1980s is thus an optimal response to the increasing globalization in the same period. We present a dynamic general equilibrium model of M&A and show that welfare, measured as aggregate consumption/production in stationary equilibrium, is a hump-shaped function of the antitrust policy parameter in the model. We are extending the model to an open economy, and aim to formalize the intuition that openness to trade demands a more lenient antitrust policy and to explore its quantitative implications for aggregate markup and welfare.

A Model of Technology Diffusion. Many new technologies, instead of being adopted simultaneously by all producers in the same area or industry, display a long and lagged diffusion process, with an S-shaped adoption curve. This even applies to technologies that were later proven to improve productivity significantly. We construct and test a model for explaining this observation. In the model, agents who are heterogeneous in beliefs choose their optimal stopping/adopting time, while they are learning from the output of others. As the population of agents who are experimenting with the new technology grows up, the learning process accelerates. Part of the incentives for them to wait is to free-ride on a larger experimenting group in the future. Our model can explain various technology diffusion data, such as the hybrid corn adoption in the U.S., or the adoption of the 12 industrial innovations.

Chapter 1

Nurturing Young Public Firms over Real Business Cycles

Linyi Cao[1]

1.1 Introduction

The study of financial intermediation is an on-going research topic existed for decades. Real-life examples of financial intermediaries, such as commercial banks, investment banks and all kinds of investment funds, for better or for worse, have proven their importance in our economy. It is essential for us to understand the roles of those financial intermediaries, their impact on the economy, and what are the consequences if we put certain regulations on them.

In this paper, I want to study financial intermediation, especially its interactions with households and entrepreneurs over real business cycles. I focus on financial intermediaries who help direct funds from households to startup firms, by performing screening on en-

[1]Department of Economics, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130 (email: *l.cao@wustl.edu*).

trepreneurs of unknown abilities. Once funded, those firms have a chance to grow large, get public, and have substantial influences on the aggregate economy. I call those financial intermediaries "nurturers of young public firms".

I want to answer two major questions. Firstly, what are the cyclical patterns of those financial intermediaries' behaviors? By behaviors I mean investment size, effort in screening entrepreneurs, efficiency in utilizing resources, etc. To answer that, I build a dynamic stochastic general equilibrium model in which the financial intermediary reacts to productivity shocks in the real sector and the financial sector promptly. Secondly, how do the cyclical behaviors reshape business cycles? In order to answer this question, I take the theory to the data, and use counterfactual exercises to evaluate the impact of financial intermediation on economic volatility and welfare.

To establish connections between my theory and the existing literature, it is necessary to briefly go through the literature of financial intermediation. Based on the natural attribute, some early researchers, such as Gurley and Shaw (1962), Benston and Smith (1976), tended to think of financial intermediaries as providing money transaction services. This is indeed a major task for commercial banks since they started to exist. But the relative importance of this task has been decreasing, even within commercial banks.

Others thought of financial intermediation as a remedy for the asymmetric information problems between investors and firms in the economy. Financial intermediaries generate social surplus by helping better allocate resources. As noted by Raymond W. Goldsmith (1969, p.400), financial intermediation "improves economic performance to the extent that it facilitates the migration of funds to the best users, i.e., to the place in the economic system where the funds will earn the highest social return".

Based on the frictions a financial intermediary solves when interacting with its counterparties, it can be performing ex-ante screening or ex-post monitoring. In the early work of Leland-Pyle (1976), Diamond (1984), financial intermediaries were modeled as solving ex-ante asymmetric information problem between lenders and borrowers. The limitation is that they were partial equilibrium analysis, so it is unlikely to say anything about the impact of financial intermediation on the economic fundamentals.

In Baron (1982), Ritter (2003), financial intermediaries were modeled to solve the asymmetric information problem faced by the firms about the financial market conditions. Firms are willing to pursue intermediaries with a good reputation, even if it means that they need to pay more. However, in those works the household side is often missing. It either appears exogenous as a stochastic demand for goods, or a stochastic supply of capital. In this paper, I'm following the classical ex-ante screening approach, but in an innovative way. I have an explicitly modeled household part, who interacts with the financial intermediary actively on different markets.

Under a grand scope, I want to develop a theory in which the following three parties and their interactions are explicitly modeled: Households who want to invest their savings but face uncertainty about the investment opportunities. Firms who need funds but face uncertainty about the fundamentals in the aggregate financial market. Financial intermediaries who interact with both parties and solve their asymmetric information problems. Financial intermediation would have a real impact on economic performance, that is, long-term growth or short-term fluctuations. An important question one could ask is whether financial intermediation dampens the economic volatility by mitigating productivity shocks, that is, a dampening effect, or the opposite, an amplification effect.

Following Galetovic (1996), Blackbur and Hung (1998), I model financial intermediation as a channel through which the economy funds firm innovations. In my model, entrepreneurs need resources to industrialize their ideas. Once funded, they have a chance to create a new variety of intermediate goods, which promises a flow of future profits. However, entrepreneurs are heterogeneous in abilities, which determine the probability of successfully creating a new variety of intermediate goods. The financial intermediary borrows resources from the households, performs screening on the entrepreneurs, and allocates resources to them.

Screening is costly to the financial intermediary, so does borrowing resources from the households. To maximize profit, the financial intermediary reacts optimally to the economic fundamentals. What I care most about are the effort intensity, measured by effort in screening per unit of investment, and the efficiency of intermediation, measure by firms created per unit of investment. If those variables are countercyclical, we could expect a dampening effect to appear, which is rarely achievable in general equilibrium models regarding financial intermediation. The intuition is that in bad (good) times, the financial intermediary is more (less) efficient in producing new varieties of intermediate goods, which mitigates the effect of productivity shocks.

It's worth mentioning that the amplification effect is somewhat prevailing in the literature studying financial stability. Diamond and Dybvig (1983) talked about the fragility of the banking system, and how it causes panic to spread during a crisis. Bernanke and Gertler (1989), Kiyotaki and Moore (1997) discussed how the worsening of credit conditions could amplify negative productivity shocks originated in the real sector. Veldkamp and Wolfers (2007) talked about coordination in information acquisition, and how it reshapes business cycles. They emphasized on how an overly-grown financial sector could undermine economic stability. My paper, on the other hand, is one of the few who try to shed light on the bright side. I want to examine whether a properly regulated financial intermediation market, who responds to business cycles promptly, can improve economic stability by reducing volatility of the fundamentals. Another difference between their works and mine is that their works are mainly on crisis, while mine is on regular business cycles.

In the theory part of the paper, I have established conditions under which there is a negative correlation between the effort intensity and the investment size, especially when the financial intermediary is flooded by resources. As a matter of fact, there exists an endogenously determined threshold value, above which the financial intermediary's optimal effort level would stop reacting to changes in its investment size. As in most macroeconomic models, investment size is procyclical, this negative correlation promises a countercyclical effort intensity, as well as the intermediation efficiency. I'm able to get these interesting qualitative results under a fairly clear and general setup. Another advantage of the model is that it also predicts a countercyclical commission rate of the financial intermediary, which is something I can test empirically.

To provide empirical evidence for my theory, I have done some works regarding the cyclicality of the commission rate. I use the U.S. IPO market as a representative market, to construct measures for investment size and commission rate. I choose the IPO market because, in a typical IPO process, the issuing firm, the investment banker(s), and the buyer(s) either in the primary market or the follow-up secondary market consist of a perfect example for the theory I have developed.

The data source is SDC: Global New Issues Database. I calculate the quarterly average commission rate of investment bankers in IPO, from 1976Q1-2016Q4. Together with economic fundamentals calculated from the U.S. economy, I have documented two main empirical facts: 1. The quarterly number and total value of IPOs are procyclical; 2. The quarterly average commission rate is countercyclical, both consistent with the model's prediction. As a robustness check, I run regressions of the commission rate on economic fundamentals controlling for size, on both the quarterly average level and the individual stock level. The countercyclicality of the commission rate remains significant.

I calibrate the model to the U.S. economy. I follow the standard real business cycle literature when calibrating parameters of the real sector and household part in the model. I calibrate parameters of the financial sector to match moments and correlations found in my empirical study of the U.S. IPO market, using the simulated method of moments. I'm not using the correlation between the cyclical output and commission rate as a target in calibration, because I want to use it to check the wellness of fit. With the calibrated model, I have simulated the cyclical output, consumption, investment size, as well as the commission rate. The correlation between the simulated cyclical output and commission rate fits the correlation found in the data pretty well: -0.15 in simulation comparing to -0.21 in the data.

I conduct several counterfactual exercises. The first exercise is to raise (reduce) the financial intermediary's cost of effort by 20%. The output volatility[2] increases (decreases) by 0.14% (0.24%). The household consumption volatility increases (decreases) by 0.32% (0.53%). The households' welfare loss (gain) is 2.22% (3.84%).[3]

In the second counterfactual exercise, I put a binding[4] commission rate cap on the financial intermediary. The output volatility increases by 0.36%. The household consumption volatility increases by 0.53%. The households' welfare loss is 5.62%. I argue that a better way to achieve a lower average commission rate is to reduce the market power of the financial intermediary. In the third exercise I shut down the financial sector completely. The output volatility increases by 0.74%. The household consumption volatility increases by 1.55%. The households' welfare loss is 7.03%.

The results from the counterfactual exercises demonstrate that in my model, financial intermediation could generate a dampening effect strong enough to overcome the typical amplification effect, and stabilize the economy.

To summarize, this paper contributes to the literature in at least three ways. In the theory part, the paper has provided an innovative mechanism under which the effort intensity and the intermediation efficiency of the financial intermediary are countercyclical, which promise a dampening effect of financial intermediation on productivity shocks. In the empirical part, the paper is one of the first to document the countercyclicality of the commission rate of financial intermediaries. In the quantitative part, the paper has demonstrated that the dampening effect is strong under a calibration which fits the data moments pretty well. The

[2] Measured by the standard deviation to mean ratio, aka the coefficient of variation.
[3] Around 90% is due to changes in the steady-state level of consumption.
[4] Much lower than the steady-state commission rate.

paper also provides normative suggestions on how to achieve a low average commission rate.

The rest of the chapter is structured as follows. Section 1.2 lays out the model and provides qualitative results. Section 1.3 describes the data and empirical evidence. Section 1.4 discusses calibration and quantitative results. Section 1.5 Concludes.

1.2 Model

To better understand the role of financial intermediation as a nurturer of young public firms, and the interactions between households, financial intermediary, and entrepreneurs over real business cycles in a stochastic general equilibrium environment, I develop an infinite horizon discrete time model with homogeneous households, a financial intermediary, heterogeneous entrepreneurs, and aggregate uncertainty.

The households face standard consumption-saving trade-off. What's new is that they also face an investment portfolio choice when they save. They can invest to replenish the capital stock, change their holdings of firms shares, or lend to the financial intermediary in an intratemporal borrowing and lending market. The financial intermediary uses resources borrowed from households to fund entrepreneurs. A funded entrepreneur can industrialize his idea to create a new variety of intermediate goods, as well as a monopolistic firm producing it, which promises a future flow of profits.

The success rate of industrialization depends on the entrepreneur's ability. But an individual entrepreneur's ability is unknown to all unless it is verified. The financial intermediary performs screening and sorting on entrepreneurs, allocates resources, and sells the successfully created intermediate goods firms to the households on a competitive market of firms shares. The surplus is split into return for the households and commission fee for the financial intermediary, either in a competitive market or through Nash bargaining, which are equivalent if allowing a taxation on the financial intermediary's profit. The aggregate uncertainty is introduced through exogenous productivity shocks.

1.2.1 Final Goods

Time is discrete and infinite $t = 0, 1, 2, \cdots$. At each period, final goods are produced competitively using intermediate goods and the following generalized constant elasticity of substitution production function

$$Y_t = \Big(A_t\Big)^{\tau - \frac{1}{\sigma-1}} \left(\int_0^{A_t} y_t(\omega)^{\frac{\sigma-1}{\sigma}} d\omega \right)^{\frac{\sigma}{\sigma-1}}. \tag{1.1}$$

An intermediate good is indexed by its variety ω, while A_t is the measure of available intermediate goods at period t. τ governs the *love of variety* as in Dixit and Stiglitz (1977), Benassy (1996). σ governs the elasticity of substitution, and is assumed to be larger than 1. At each time t, the final goods are used as the numeraire.

Denote $p_t(\omega)$ as the price of intermediate good ω at period t, I can derive its demand function

$$p_t(\omega) = A_t^{\frac{\tau(\sigma-1)-1}{\sigma}} \left(\frac{Y_t}{y_t} \right)^{\frac{1}{\sigma}}. \tag{1.2}$$

1.2.2 Intermediate Goods

Each variety of intermediate goods is produced by a single monopolistic firm. The production of intermediate goods uses capital as the sole input and according to a concave production function

$$y_t(\omega) = \exp(\varepsilon_t) k_t^{\alpha}(\omega).$$

ε_t represents the contemporary aggregate productivity shock in the intermediate goods sector, whose law of motion will be specified in the household part later. $\alpha \in (0, 1)$ governs concavity and the capital share. An intermediate goods firm takes the demand function (1.2) and the capital interest rate r_t as given, and solves

$$\pi_t(\omega) = \max_{y_t(\omega)} p_t(\omega)y_t(\omega) - (r_t + \delta)\left[\exp(-\varepsilon_t)y_t(\omega)\right]^{\frac{1}{\alpha}},$$

where $\delta \in (0, 1)$ is the capital depreciation rate.

Solving the profit maximization problem above, I have the optimal quantity

$$y_t^*(\omega) = \left(\frac{\sigma - 1}{\sigma}\frac{\exp(\varepsilon_t/\alpha)}{r_t + \delta}\alpha A_t^{\frac{\tau(\sigma-1)-1}{\sigma}} Y_t^{\frac{1}{\sigma}}\right)^{\frac{\alpha\sigma}{\alpha+\sigma-\alpha\sigma}}. \tag{1.3}$$

And most importantly, the intermediate goods firm earns a profit $\pi_t(\omega) > 0$. I assume that the profit is paid to shareholders of the firm each period as dividend. One can see that intermediate goods firms are symmetric in this economy.

Each intermediate goods firm acts as a monopolist because it owns the patent to produce a certain variety of intermediate goods. New varieties of intermediate goods are created by entrepreneurs. Once funded, an entrepreneur has the chance to industrialize his idea to create a new variety of intermediate goods, as well as a firm producing it. The newly created firm is temporarily owned by the financial intermediary who funds the entrepreneur, and is sold to the households right after on a competitive firm's shares market, while the entrepreneur gets a one-time payment for successfully industrializing his idea. One can see a direct and intuitive connection between the process described above, and the real-life initial public offering (IPO), private equity (PE), or investment banking. I will be more specific about this point in the latter entrepreneur and financial intermediary part.

After finishing the current period production, all incumbent intermediate goods firms face an exogenous death rate of $\eta \in (0, 1)$. I define Q_t as the post-dividend expected present value of an intermediate goods firm that survives period t. It is written as[5]

$$
\begin{aligned}
Q_t \equiv \quad & \mathbb{E}_t\left[\sum_{\tau=1}^{\infty}(1-\eta)^{\tau-1}\frac{\beta^\tau u'(c_{t+\tau})}{u'(c_t)}\pi_{t+\tau}\right] \\
= \quad & \mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)}\Big(\pi_{t+1}+(1-\eta)Q_{t+1}\Big)\right],
\end{aligned}
\tag{1.4}
$$

which is derived in equation (1.20). Q_t is a crucial endogenous variable for the financial market, because it is essentially the price of a newly created intermediate goods firm in period t. Since in this economy a new intermediate goods firm is equivalent to a new variety of intermediate goods, I will use these two terms interchangeably later.

Now that I have described the goods production sector, it's time to look at the three main parties of agents in this economy. They are, entrepreneurs who have ideas and need resources. Households who have resources and try to find users promising highest returns. The financial intermediary who has access to a screening technology and generate social surplus by helping better allocate resources to entrepreneurs. Figure 1.1 illustrates the interactions between the three parties in the economy.

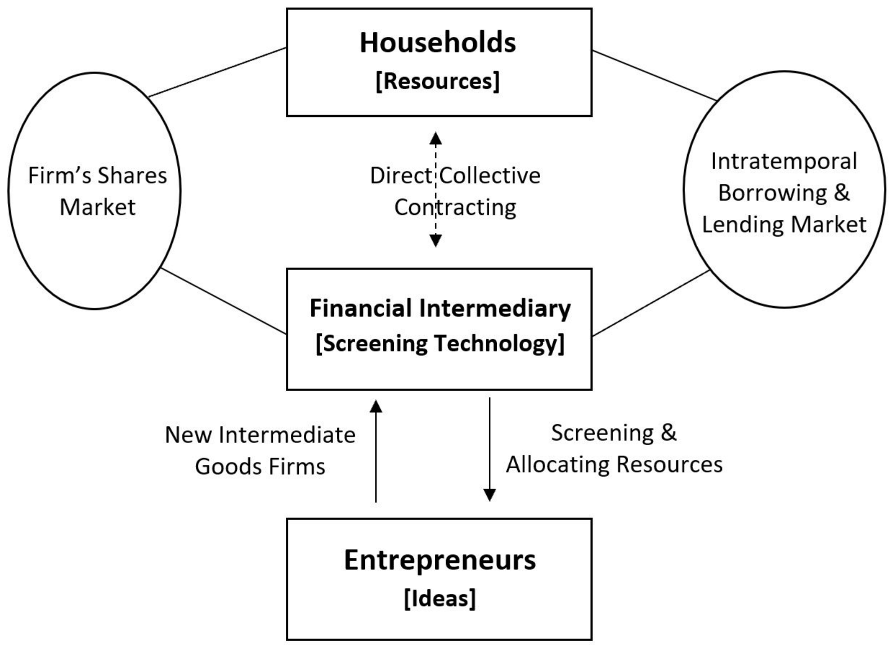
1.2.3 Entrepreneurs

The economy is populated by a large[6] continuum of one-period lived, risk-neutral entrepreneurs, each of whom is endowed with an idea that might be industrialized, into a new variety of intermediate goods. Entrepreneurs require $\xi > 0$ amount of final goods to industrialize the idea, but are heterogeneous in their abilities $z \in [0, 1]$. That is, an entrepreneur with ability z, upon spending ξ units of final goods, will have a probability z to successfully create a

[5]The formula shall be derived from the individual household's optimization problem. Since households in my model are homogeneous, I use c_t to denote the per capita consumption at time t, which is common across households in equilibrium.

[6]By large I mean it has the same measure as the real line $\mathbb{R}$, so the pool of entrepreneurs (ideas) will never be exhausted by the financial intermediary. This point will be more clear in the financial intermediary part.

Figure 1.1: Interactions Between Agents in the Economy



new variety of intermediate goods. To be consistent with the pricing formula (1.4), I assume that an intermediate goods firm created in period t survives the current period for sure, and starts to produce in period $t+1$. I will drop the subscript t from now on, because it is irrelevant in the rest of this section.

The measure of entrepreneurs is assumed to be large, comparing to the measure of households, which is normalized to 1. I make this assumption because I believe ideas are plenty, but the resources needed to make them realize are always constrained. The cost of industrialization ξ should be thought of as an amount of resources no individual household could afford alone. Also notice that ξ is exogenous and fixed across entrepreneurs, while in the data we do observe entrepreneurs of different scope asking for a different amount of investment, and end up creating firms of various sizes. This assumption is a solid one in my model though, for two reasons. First is that in the model intermediate goods firms are symmetric, so there is no room for "dream bigger, ask for more". Second is that I can bring in heterogeneous demand for investment even under this framework, by allowing one entrepreneur having multiple ideas. However, it complicates the model while bringing about little new intuition.

The creation of a new intermediate goods firm is publicly observable, but I assume that a particular entrepreneur's ability z is unknown to all, including the entrepreneur himself, unless verified by a financial intermediary. This serves as the main friction in the economy. The reservation utility of the entrepreneurs is assumed to be time-invariant and normalized to 0.[7] The population distribution of entrepreneur abilities is characterized by a cumulative distribution function $H(z)$, which is common knowledge. Denote the population mean of abilities as $\mathbb{E}(z) \equiv \int_0^1 z \, dH(z)$.

Assumption 1. *The distribution of abilities $H(z)$ has a density $h(z)$ which satisfies*

$$h(z) > 0, \ \forall z \in [0, 1].$$

This assumption is to make sure that $H(z)$ has a well-defined inverse function on $[0, 1]$, and both $H(z)$ and its inverse function are differentiable.

1.2.4 Financial Intermediary

The economy has one[8] risk-neutral financial intermediary, who is assumed to be impatient.[9] The financial intermediary solves a static profit maximization problem every period. It has access to a screening technology.[10] Once screened by the financial intermediary, an

[7] I can assume this because the entrepreneurs are not aware of their own abilities, so they are ex-ante homogeneous.

[8] One could think of it as a unity of all financial intermediaries, or the financial sector.

[9] This assumption is to rule out intratemporal decisions of the financial intermediary such as hoarding. One can think of the financial intermediary as being replaced by a new one every period.

[10] I'm not assuming that an individual household can not operate such technology. But because of the economy of scale, the financial intermediary could operate the technology more efficiently.

entrepreneur's ability z becomes known to the financial intermediary and the entrepreneur himself, but not to any other agents in the economy.

Every period, upon putting in effort level $e \in \mathbb{R}_+$, the financial intermediary screens a measure e of entrepreneurs, randomly drawn from the population. The population of entrepreneurs (ideas) is assumed to be so large that it can not be exhausted by the financial intermediary.[11] Screening gives the financial intermediary a pool of verified entrepreneurs, among whom the exact mapping between individuals and abilities are known by the financial intermediary. Sorting is thus automatic within the pool of verified entrepreneurs. Those who are left out consist of a pool of unverified entrepreneurs, among whom the mapping between individuals and abilities remains unknown. Since e is a positive measure and drawing is random, both the pool of verified entrepreneurs and the pool of unverified ones should inherit the original population distribution of abilities $H(z)$. This turns out to be an important property later in determining the optimal allocation rule of resources.

The screening effort is observed only by the financial intermediary, and incurs a private cost $F(e)$ in unit of final goods.

Assumption 2. *The cost function $F(e)$ has the following properties:*

(I) $F(0) = 0,\ F'(0) = 0$;

(II) $F'(e) > 0,\ F''(e) > 0,\ \forall e \in \mathbb{R}_{++}$;

(III) $F(e)$ may have a small fixed cost component.

Basically, I want the cost of screening effort to be strictly increasing and strictly convex. The requirement for a very small fixed cost component is to guarantee an interior solution of the optimal effort level. It is important to point out that screening is costly, but randomly drawing entrepreneurs from the population is not. The financial intermediary can draw cost-freely as many entrepreneurs as it wants at any time, without screening them.

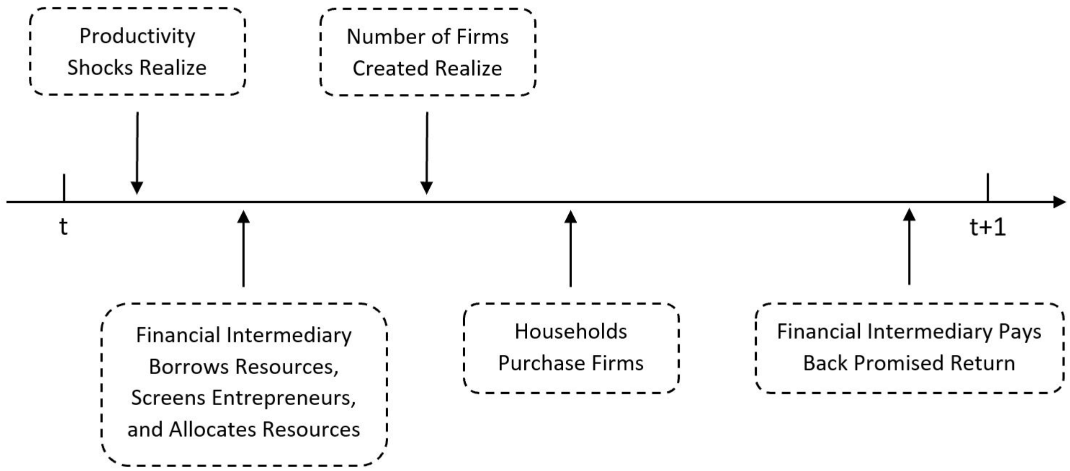
[11] Basically, the financial intermediary can draw entrepreneurs with replacement.

As stated in the intermediate goods production part, a variety of intermediate goods defines a monopolistic firm. And the expected present value of a new intermediate goods firm is Q. Remember that in this economy the entrepreneurs and the financial intermediary are assumed to be one-period lived and myopic respectively, they don't own any resources or hold any firm's shares at the beginning of a period. The households are the ultimate provider of resources, and the only candidate for purchasing and holding intermediate goods firms. Every period, after observing the productivity shocks, the financial intermediary borrows X amount of final goods from the households, with a promised intratemporal return rate R, in an intratemporal borrowing and lending market that is either competitive or centralized. It then performs screening, and allocates resources to entrepreneurs. After the new intermediate goods firms are created, the financial intermediary sells the ownership to the households on a competitive firm's shares market, and pay the promised return to the households. The commission fee would be the difference between the value of the firms and the return promised to the households. The rest of the section is to characterize how the amount of intermediated investment X and the screening effort level e are determined.

The result that the intermediate goods firms are owned by various households is consistent with what's happening in reality. In the U.S., the shareholders of public firms are on average highly diversified, comparing to startup firms before going public. I would think of those public firms as intermediate goods firms in the model, while those small private firms as ideas, solely owned by one or a few persons.

Let's start with the simplest piece of the puzzle, the entrepreneurs. The entrepreneurs are rather passive in this economy. They can be paid at different times, but nevertheless getting paid the reservation utility. To see this, think of an entrepreneur before being screened. He is definitely getting paid the reservation utility because there are abundant of them in the economy. Now think of an entrepreneur who has already been screened by the financial intermediary, his outside option doesn't change because this piece of information about his

Figure 1.2: A Timeline of Events for the Financial Intermediary



ability is not shared with any third party. The financial intermediary only needs to pay the entrepreneurs enough to induce participation.

This setup of entrepreneurs allows for interesting extensions though. For example, imagine that there exists a moral hazard problem between the financial intermediary and the entrepreneurs. After receiving ξ amount of final goods from the financial intermediary, an entrepreneur could use it to industrialize idea as described before, or to divert it into personal consumption, which yields him a utility of $\phi\xi$. The financial intermediary observes the outcome of industrialization, but can't directly control the action taken by the entrepreneur. An optimal contract thus arises between the financial intermediary and a funded entrepreneur whose ability is z: If the industrialization is successful, the entrepreneur gets paid $\frac{\phi\xi}{z}$, otherwise 0. The contract would make it incentive compatible for the entrepreneurs to use the resources for industrialization. This simple extension is equivalent to a setup in which the ex-ante reservation utility is $\phi\xi$ rather than 0.

The real interesting interaction arises between the households and the financial intermediary. At each period, there is an intratemporal borrowing and lending market, decentralized or centralized, between the households and the financial intermediary. The financial intermediary makes three decisions, first it decides how much resources X to borrow from the households. Then it decides how much effort e to put in screening entrepreneurs. At last, it decides how to allocate resources to the entrepreneurs, verified or unverified. In funding entrepreneurs, I assume the following concave technology

Assumption 3. *With X unit of resources, the financial intermediary can fund a maximum number of $N = exp(\chi)\dfrac{n(X)}{\xi}$ entrepreneurs, where χ represents the aggregate productivity shock in the financial sector. $n(X)$ is continuously differentiable and satisfies*

(I) $0 \leq n(X) \leq X$, $\forall X \in [1, \infty)$;

(II) $n'(X) > 0$, $n''(X) < 0$, $\forall X \in \mathbb{R}_{++}$.

The assumption implies that there is efficiency loss as the size of intermediated investment X gets large. This efficiency loss can be rationalized by a higher cost of bookkeeping, or as a reduced representation of a higher cost of monitoring entrepreneurs. The introduction of an aggregate productivity shock χ in the financial sector is for quantitative purposes. One can also regard it as a reduced representation of variations in the cost of industrialization: $\exp(-\chi)\xi$. Given χ, there is a one-to-one mapping between how much resources to be allocated X and how many entrepreneurs to be funded N. Since the financial intermediary makes decisions after observing the productivity shocks, later on I will treat the decision of X and the decision of N as equivalent.

Now that I have described the technologies available for the financial intermediary, it's time to specify the market structure in the financial sector, that is, the firm's shares market and the intratemporal borrowing and lending market. I assume that the firm's shares market is competitive. The financial intermediary is a price-taker even though it supplies a positive measure of firms.[12] As for the intratemporal borrowing and lending market, I assume that it also operates competitively each period, for the financial intermediary to borrow resources

[12] This assumption is to rule out strategic behaviors of the financial intermediary when supplying newly created firms to the market.

from the households. The return rate R serves as the price. Since it is an intratemporal market with no other frictions, the households' supply of resources is perfectly elastic.[13]

The financial intermediary takes the price of firms Q and the return rate to households R as given, and maximizes its profit (consumption) after observing the productivity shocks. I denote the number of firms created as S, and give the maximization problem

$$\max_{X,e,\text{allocation rule}} Q \bullet S(X, e, \text{allocation rule}) - F(e) - R \bullet X. \tag{1.5}$$

To solve the problem, I break it into three steps, one for each decision the financial intermediary has to make. Let's start with the resource allocation rule. For any given size of entrepreneurs to be funded N[14] and screening effort level e, the financial intermediary would want to choose an allocation rule that maximizes the number of new firms created. The space of all feasible allocation rules is huge, and can not be summarized by a finite number of variables. Fortunately, each allocation rule maps into a distribution of abilities of those entrepreneurs who are funded, which I can use to calculate the number of firms created. Moreover, I'm going to argue that the optimal allocation rule is essentially a threshold rule. The financial intermediary first utilize the pool of verified entrepreneurs. It should start with verified entrepreneurs of the highest ability, then gradually go down as those are exhausted, until reaching the natural lower bound $\mathbb{E}(z)$. If there were any resources left, the financial intermediary should tend to the pool of unverified entrepreneurs, who promises an average success rate of $\mathbb{E}(z)$. That is, entrepreneurs whose ability is verified to be worse than the population mean shall never get funded, otherwise, they get funded in a descending order.

It is the $\frac{e}{N}$ ratio that determines who gets funded, which is also a measure of "effort

[13]That is, when the return rate is less (more) than 1, an individual household supplies resources of quantity 0 (up to its income). When the return rate is 1, the household is indifferent between how much resources to supply.

[14]I'm solving the optimal allocation rule for any given N instead of X because firstly they are equivalent, secondly using N gives me neat analytical results.

intensity". To understand why the population mean $\mathbb{E}(z)$ serves as the natural lower bound, remember when screening entrepreneurs, the drawing is random. So both the verified pool and the unverified pool inherit the original population distribution of abilities $H(z)$. To summarize the allocation rule I have just described in a more rigorous way, I give the following definition.

Definition 1. *(The Descending Allocation Rule)*

Suppose the $\frac{e}{N}$ *ratio is large enough such that* $e\big[1 - H\big(\mathbb{E}(z)\big)\big] \geq N$*, only the pool of verified entrepreneurs is in use. Verified entrepreneurs whose ability is above threshold* $\underline{z}$ *shall be funded, where* $\underline{z}$ *solves*

$$e\big[1 - H(\underline{z})\big] = N. \tag{1.6}$$

Otherwise, both the pools of verified and unverified entrepreneurs are in use. Verified entrepreneurs whose ability is above $\mathbb{E}(z)$ *shall be funded, the remaining resources go to unverified, randomly drawn entrepreneurs.*

With the explanation above, the proof of why this descending allocation rule is the optimal allocate rule which maximizes the number of firms created for any given N and e should be straightforward. The threshold $\underline{z}$ is a decreasing function of N and an increasing function of e, which has a closed-form representation

$$\underline{z}(N, e) = H^{-1}\bigg(1 - \frac{N}{e}\bigg).$$

The optimal allocation rule is essentially a threshold rule among the verified entrepreneurs, but with the twist that some resources might also go to unverified entrepreneurs. The effective ability threshold is given by

$$\hat{z}(N, e) = max\Big\{\underline{z}(N, e), \mathbb{E}(z)\Big\}.$$

Following the descending allocation rule described in Definition 1, I can derive the number of firms created as

$$S(N,e)=\begin{cases} e\displaystyle\int_{\mathbb{E}(z)}^{1}\big(z-\mathbb{E}(z)\big)h(z)dz+N\bullet\mathbb{E}(z), & \dfrac{e}{N}<\dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \\ e\displaystyle\int_{\underline{z}(N,e)}^{1}zh(z)dz, & \dfrac{e}{N}\geq\dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \end{cases}, \tag{1.7}$$

which I call the "success function". Detailed derivation can be found in Appendix 1.6.1.

Several things to notice about $S(N,e)$. It is continuous in N and e but demonstrates two regions. When e is small, the verified entrepreneurs whose ability is above the population mean won't be enough to exhaust all the resources. Part of the resources will be allocated to randomly drawn, unverified entrepreneurs. In that case, the benefit of a higher effort level is on the extensive margin, that is, a reallocation of resources from unverified entrepreneurs back to the verified and qualified ones. This benefit shall be constant and large. However, when e is large enough, verified and qualified entrepreneurs will exhaust all the funds. So the benefit is on the intensive margin, namely from an improved reallocation among those verified and qualified entrepreneurs. And this benefit shall be diminishing. To be more specific about the shape of $S(N,e)$, I give the following proposition.

Proposition 1. *Under Assumption 1, the success function $S(N,e)$ is continuously differentiable, strictly increasing, and concave on $\mathbb{R}^2_+$. Moreover, it is linear when $\dfrac{e}{N}<\dfrac{1}{1-H\big(\mathbb{E}(z)\big)}$ and strictly concave when $\dfrac{e}{N}>\dfrac{1}{1-H\big(\mathbb{E}(z)\big)}$.*[15]

The fact that $S(N,e)$ is strictly increasing in e makes it a perfect signal of the otherwise hidden effort level. The curvature of $S(N,e)$ implies a constant marginal benefit of effort when the screening effort e is below cutoff $\dfrac{N}{1-H\big(\mathbb{E}(z)\big)}$, and a decreasing marginal benefit of effort when above. In addition, a change of N moves the effort cutoff around, but has

[15] Being linear means $S\big(\varphi N_1+(1-\varphi)N_2,\varphi e_1+(1-\varphi)e_2\big)=\varphi S(N_1,e_1)+(1-\varphi)S(N_2,e_2),\ \forall\varphi\in(0,1)$. Being strictly concave means $S\big(\varphi N_1+(1-\varphi)N_2,\varphi e_1+(1-\varphi)e_2\big)>\varphi S(N_1,e_1)+(1-\varphi)S(N_2,e_2),\ \forall\varphi\in(0,1)$.

different effects on the marginal benefit of effort below or above the cutoff. To summarize rigorously, I given the following corollary.

Corollary 2. *The marginal benefit of effort, measured by $Q \bullet S_e(N, e)$, has the following properties:*

(I) The marginal benefit of effort is bounded above by $Q \int_{\mathbb{E}(z)}^{1} \big(z - \mathbb{E}(z)\big) h(z) dz$, and is weakly decreasing in e. To be more specific, it is constantly at the upper bound when e is below the cutoff, and is strictly decreasing once e exceeds the cutoff;

(II) An increase of N does not affect the marginal benefit of effort on its constant region, but enlarges the constant region. It shifts up the marginal benefit of effort on the decreasing region;

(III) An increase of Q shifts up the marginal benefit of effort proportionally everywhere.

The proof of Proposition 1 and Corollary 2 is given in Appendix 1.6.2. These properties about the marginal benefit of effort, $Q \bullet S_e(N, e)$, are crucial when determining the optimal effort level. They also carry the intuition of why e does not respond to N if it is too large.

So far I have introduced the optimal rule of allocating resources to the entrepreneurs for any given funded entrepreneurs' size N, or equivalently, intermediated investment size X, and effort level e. I have also established the properties of the success function $S(N, e)$, which measures the number of firms created under the optimal allocation rule. I give the following profit maximization problem, which is updated from the original profit maximization problem (1.5) using the optimal allocation rule.

$$\underset{X,e}{max}\ Q \bullet S(N, e) - F(e) - R \bullet X \tag{1.5'}$$

s.t.

$$N = \exp(\chi) \frac{n(X)}{\xi}.$$

In Proposition 1, I have established the concavity of the $S(N,e)$ function. The objective function in maximization problem (1.5') is strictly concave as long as $F(e)$ is strictly convex and $n(X)$ is strictly concave, which are guaranteed by Assumption 2 and 3 respectively. The maximization problem yields a unique solution of the optimal borrowing size X^* and the optimal effort level e^*, which are characterized by the following first order conditions

$$Q \bullet S_N(N,e) \bullet \exp(\chi)\frac{n'(X)}{\xi} = R. \tag{1.8}$$

$$Q \bullet S_e(N,e) = F'(e), \tag{1.9}$$

I have the optimal size of entrepreneurs to be funded as $N^* = \exp(\chi)\frac{n(X^*)}{\xi}$. The number of firms created is

$$S^* = S(N^*, e^*). \tag{1.10}$$

The commission fee for the financial intermediary is

$$M^* = Q \bullet S^* - R \bullet X^*, \tag{1.11}$$

while the profit (consumption) of the financial intermediary is

$$M^* - F(e^*).$$

As has been discussed in the introduction part of the paper, what I care most about in the financial market are three ratios. First is the effort-to-size ratio, measured either by $\frac{e^*}{X^*}$ or by $\frac{e^*}{N^*}$. I refer to it as the "effort intensity", and this ratio measures how hard the financial intermediary works per unit of investment. Second is the success-to-size ratio, measured either by $\frac{S^*}{X^*}$ or by $\frac{S^*}{N^*}$. I refer to it as the "average ability of funded entrepreneurs", or as the "average quality of financial services". This ratio measures the financial intermediary's efficiency in turning input (final goods) into output (intermediate goods firms). The last one is the commission-to-value ratio, measured by $\frac{M^*}{Q \bullet S^*}$. And it is the "commission rate" which we could observe from the data.[16]

From the first order conditions (1.8) and (1.9), one can see the optimal borrowing size and effort level are functions of Q and χ. Using simple comparative statics, one can check that $X^*(Q,\chi)$ is strictly increasing in Q and χ, while $e^*(Q,\chi)$ is strictly increasing in Q but not necessarily in χ. Generally speaking, X^* and e^* are moved around by Q and χ across the cycles. With different values of Q and realizations of χ, there could be two scenarios. One is the optimal X^* and e^* are such that

$$\frac{e^*}{N^*} < \frac{1}{1 - H\big(\mathbb{E}(z)\big)}. \tag{1.12}$$

I'm particularly interested in this scenario because it lays the foundation of achieving a negative correlation between the optimal effort intensity $\frac{e^*}{X^*}$ and the borrowing size X^*. More specifically, I'm going to show that if in equilibrium (1.12) is met, $e^*(Q,\chi) = e^*(Q)$.

To take a closer look at the scenario above, I further break the maximization problem (1.5') into two steps. First, the financial intermediary chooses the effort level e while taking N as given, that is

$$\max_e \; Q \bullet S(N,e) - F(e),$$

which yields an effort policy function $e(N)$. Then it chooses the optimal borrowing size X^*, taking into consideration of its effect on $e(N)$, that is

$$\max_X \; Q \bullet S\big(N, e(N)\big) - F\big(e(N)\big) - R \bullet X \tag{1.5''}$$

[16] An alternative measure of commission rate is the commission-to-size ratio $\frac{M^*}{X^*}$. Using (1.11) one can easily prove that $\frac{M^*}{X^*}$ and $\frac{M^*}{Q \bullet S^*}$ have the same cyclicality.

s.t.

$$N = \exp(\chi)\frac{n(X)}{\xi}.$$

The eventual optimal effort level would be $e^* = e(N^*)$.

The properties of the $e(N)$ function are crucial for understanding why when X (or equivalently N) is large, the effort intensity necessarily drops. I give the following theorem

Theorem 3. *Under Assumption 1 and 2, for any given Q, there exists a threshold value $\bar{N}(Q)$ such that $e(N)$ intersects with $\frac{N}{1-H\big(\mathbb{E}(z)\big)}$ at $\bar{N}(Q)$ and*

When $N < \bar{N}(Q)$, $e(N)$ is strictly increasing in N and $e(N) > \frac{N}{1-H\big(\mathbb{E}(z)\big)}$;

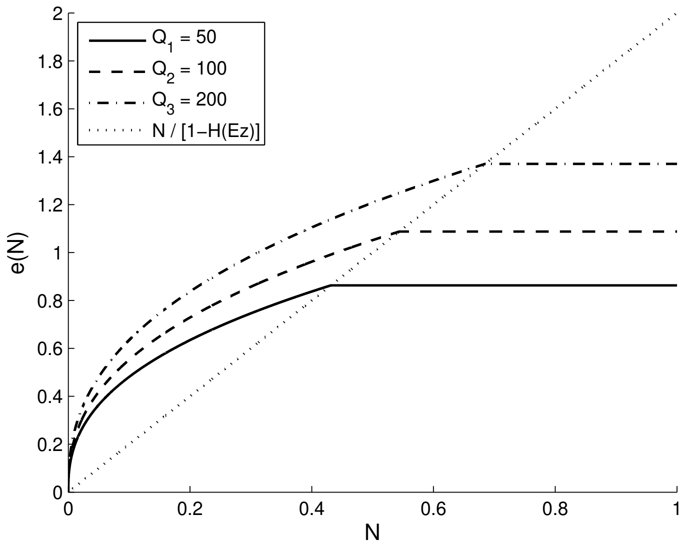
When $N > \bar{N}(Q)$, $e(N)$ is constant in N and $e(N) < \frac{N}{1-H\big(\mathbb{E}(z)\big)}$.

In addition, the threshold value $\bar{N}(Q)$ is strictly increasing in Q.

The proof of Theorem 3 is given in Appendix 1.6.3. Much to be said about this effort policy function. The function is continuous but with kink points at $N = \bar{N}(Q)$. And it is bounded above when N increases, which indicates that a larger size of investment won't necessarily induce a higher level of screening effort. As a matter of fact, when flooded by resources, the financial intermediary will only screen a fixed amount of entrepreneurs, and allocate the excessive resources to randomly drawn, unverified ones. More specifically, When $N < \bar{N}(Q)$, the intersection of $Q \cdot S_e(N, e)$ and $F'(e)$ must yield an effort level e above the effort cutoff, so only the verified entrepreneurs are funded. But once $N > \bar{N}(Q)$, the intersection will yield an e below the effort cutoff, and part of the resources goes to unverified entrepreneurs. I call $\big(\bar{N}(Q), \infty\big)$ the "excess region" of N, and the corresponding $\big(\bar{X}(Q), \infty\big)$ the "excess region" of X.

From the first order conditions (1.8) and (1.9), one can see that χ affects the optimal effort level e^* only if N affects (1.9). Based on Theorem 3, I conclude if the optimal borrowing size X^* is such that $N^* > \bar{N}$, the optimal effort level $e^* = e(N^*)$ is constant w.r.t. χ, and

Figure 1.3: The $e(N)$ Function with Different Values of Q



we are in the scenario specified by (1.12). An increase of X^* or N^* due to χ will necessarily reduce the effort intensity. This sheds light on the countercyclicality of the effort intensity, as well as the average quality of financial services and the commission rate. In data, as well as predicted by my model, the size of aggregate investment is highly procyclical. In booms (recessions), households invest more (less) in general, they also invest more (less) through the financial intermediary. So a negative correlation between the investment size and the effort intensity is essential for achieving the countercyclicality of those variables. On the excess region, the negative correlation holds for sure. Below the excess region, however, the correlation between $\frac{e^*}{X^*}$ and X^* remains ambiguous. So we should expect the countercyclical pattern to be more prominent during booms, cause X^* is more likely to stay in the excess region.

I have proved the existence of an excess region, and established the fact that the optimal effort level doesn't vary with the financial sector productivity shock when the optimal borrowing size lies in that region. However, both e^* and X^* are still moved around by the price of firms Q across the cycles. The intuition behind is straightforward, if the value of an intermediate goods firm gets higher, the financial intermediary should react with a larger investment size, and put more effort into screening entrepreneurs. As in most macro-asset pricing models, Q is procyclical. So to guarantee the countercyclicality of the effort intensity, the average quality financial services, and the commission rate, I need two conditions: 1. The steady-state X^*_{ss} is on its excess region; 2. $e^*(Q)$ is less sensitive to Q comparing with $X^*(Q,\chi)$, which shall be satisfied if the elasticity of $F'(e)$ is larger than that of $\frac{1}{n'(X)}$. Unfortunately, I couldn't provide a more precise characterization regarding condition 2 without further specifying some functional forms.

I have described the determination of the optimal allocation rule, borrowing size, and screening effort level, under a setup which is as general as possible. I have also elaborated on the mechanism through which a negative correlation between the effort intensity and the investment size could be generated. To provide more analytical results, I need to specify the functional forms of $F(e)$ and $n(X)$.[17]

Assumption 4. *The cost function has the following form*

$$F(e) = \mathbb{1}\{e > 0\} f^0 + \frac{f}{1+\kappa} e^{1+\kappa};\ \kappa > 0.$$

And the fixed cost component f^0 is very small.

The technology of handling resources has the following form

$$n(X) = X^{\lambda};\ \lambda \in (0,1).$$

[17]What I need is that $F'(e)$ is more elastic compared to $\frac{1}{n'(X)}$. Using the following functional forms, this property is easy to characterize.

I further require

$$\kappa > \frac{1-\lambda}{\lambda}.$$

Functional forms specified in Assumption 4 are chosen to be consistent with Assumption 2 and 3, and being easy to handle. The requirement of a large enough κ is to guarantee that $F'(e)$ more elastic than $\frac{1}{n'(X)}$. Suppose the optimal X^* and e^* are indeed such that (1.12) is met, which shall be verified later. Together with the functional forms, I can solve for

$$X^* = \left(\frac{\exp(\chi)\lambda\mathbb{E}(z)Q}{R\xi}\right)^{\frac{1}{1-\lambda}},$$

$$N^* = \frac{1}{\xi}\left(\frac{\exp(\chi)^{\frac{1}{\lambda}}\lambda\mathbb{E}(z)Q}{R\xi}\right)^{\frac{\lambda}{1-\lambda}},$$

and

$$e^* = \left(\frac{\Delta_{H(z)}Q}{f}\right)^{\frac{1}{\kappa}},$$

where $\Delta_{H(z)} \equiv \int_{\mathbb{E}(z)}^{1} \big(z - \mathbb{E}(z)\big)h(z)dz$ is a constant determined solely by the population distribution of abilities $H(z)$.

I give the condition under which (1.12) is met as

$$Q > \left[\left(\frac{\xi}{\exp(\chi)}\right)^{\kappa}\Big(1 - H\big(\mathbb{E}(z)\big)\Big)^{(1-\lambda)\kappa}\left(\frac{\Delta_{H(z)}}{f}\right)^{1-\lambda}\left(\frac{R}{\lambda\mathbb{E}(z)}\right)^{\lambda\kappa}\right]^{\frac{1}{\lambda\kappa-(1-\lambda)}}.$$

Which is more likely to happen in booms cause both Q and χ are procyclical.

Now I can check the ratios, what move them around in equilibrium are the price of firms Q and the financial sector productivity shock χ. Let's start with the effort-to-size ratio, or the "effort intensity". I derive

$$\frac{e^*}{X^*} = \left(\frac{\Delta_{H(z)}}{f}\right)^{\frac{1}{\kappa}}\left(\frac{R\xi}{\exp(\chi)\lambda\mathbb{E}(z)}\right)^{\frac{1}{1-\lambda}} Q^{\frac{1}{\kappa}-\frac{1}{1-\lambda}},$$

and

$$\frac{e^*}{N^*} = \xi \left(\frac{\Delta_{H(z)}}{f}\right)^{\frac{1}{\kappa}} \left(\frac{R\xi}{\exp(\chi)^{\frac{1}{\lambda}}\lambda\mathbb{E}(z)}\right)^{\frac{\lambda}{1-\lambda}} Q^{\frac{1}{\kappa}-\frac{\lambda}{1-\lambda}}.$$

Based on Assumption 4, I have

$$\frac{1}{\kappa} < \frac{\lambda}{1-\lambda} < \frac{1}{1-\lambda},$$

which indicates a negative correlation between $\frac{e^*}{X^*}$ and Q, so are $\frac{e^*}{N^*}$ and Q. Also notice that χ only shows up in the denominator, so $\frac{e^*}{X^*}$ and $\frac{e^*}{N^*}$ are negatively correlated with χ as well.

Now let's check the success-to-size ratio, or the "average quality of financial services". Remember that when (1.12) is met,

$$S^* = e^* \bullet \Delta_{H(z)} + N^* \bullet \mathbb{E}(z).$$

And one can check $\frac{N^*}{X^*}$ is decreasing in Q and constant in χ. Thus both $\frac{S^*}{X^*}$ and $\frac{S^*}{N^*}$ are negatively correlated with Q and χ.

The last is to check the commission-to-value ratio, or the "commission rate". I can derive

$$\frac{M^*}{Q \bullet S^*} = 1 - \frac{R}{\Delta_{H(z)}\frac{Q \bullet e^*}{X^*} + \mathbb{E}(z)\frac{Q \bullet N^*}{X^*}}.$$

And one can check $\frac{Q \bullet e^*}{X^*}$ is decreasing in Q and χ, while $\frac{Q \bullet N^*}{X^*}$ is constant in both Q and χ. That is, $\frac{M^*}{Q \bullet S^*}$ is also negatively correlated with Q and χ.

As I have mentioned previously, the price of firms is higher when the economy experiences a positive productivity shock in the intermediate goods sector. Meanwhile, a positive productivity shock in the financial sector boosts up future output. These two facts, together with the properties discussed above, promise the countercyclicality of those ratios.

So far, I have laid out the setup of the financial sector and established conditions under which the effort intensity, the average quality of financial services and the commission rate are countercyclical. However, there is a shortage to this competitive setup. That is, one can not vary the "market power" or profit of the financial intermediary easily because both the supply of resources and the demand for firm's shares are perfectly elastic. It causes problems for me to take the model to the data. That is why I have introduced an alternative centralized setup in Appendix 1.6.4, in which the market power of the financial intermediary is varied by its bargaining power γ in a Nash bargaining with the households. I have also established the equivalence between these two setups. What I need in the competitive setup is to add a taxation on the financial intermediary's profit, and use it to compensate the households with a lump-sum transfer

$$T = \underbrace{(1-\gamma)}_{\text{tax rate}} \big[Q \bullet S(N^*, e^*) - F(e^*) - R \bullet X^* \big].$$

This taxation scheme has a pure redistribution effect in the sense that it won't affect any participants' incentives in the financial sector.

1.2.5 Households

The economy is populated by a continuum of infinitely lived, risk-averse, and homogeneous households, whose measure is normalized to 1. Households own the capital stock and the stock of intermediate goods firms in the economy. Each period, an individual household's wealth consists of the rent and value from holding capital stock k_t, the pre-dividend value of holding intermediate goods firm's shares a_t, and a lump-sum transfer T_t. It then decides how much to consume c_t, how much capital stock k_{t+1} and intermediate goods firm's shares a_{t+1} to take to the next period, and how much x_t^h to lend to the financial intermediary in the intratemporal borrowing and lending market. I denote s_t^h as the net change in an individual household's holding of firm's shares.[18] The households make decisions after observing the contemporary productivity shocks ε_t, χ_t.

An individual household is a price-taker in the financial markets and the capital renting market. The household's utility from consumption is $u(c)$, satisfying $u'(c) > 0$, $u''(c) < 0$, and $\lim\limits_{c \to 0} u'(c) = \infty$, with time discount factor $\beta \in (0, 1)$. I give the individual household's optimization problem as the following.

$$\max_{\{c_t, k_{t+1}, a_{t+1}, x_t^h\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t u(c_t)$$

s.t. $\forall t$ and realization of the productivity shocks $\varepsilon_t,\ \chi_t$

$$c_t + k_{t+1} + Q_t a_{t+1} = (1 + r_t) k_t + \big[\pi_t + (1 - \eta) Q_t\big] a_t + (R_t - 1) x_t^h + T_t, \tag{1.13}$$

$$x_t^h \leq (1 + r_t) k_t + \big[\pi_t + (1 - \eta) Q_t\big] a_t, \tag{1.14}$$

$$\begin{pmatrix} \varepsilon_{t+1} \\ \chi_{t+1} \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \chi_t \end{pmatrix} + \begin{pmatrix} \nu_{t+1} \\ \mu_{t+1} \end{pmatrix}, \tag{1.15}$$

$$c_t,\ k_{t+1},\ a_{t+1} \geq 0. \tag{1.16}$$

As a reminder, ε_t represents the contemporary productivity shock in the intermediate goods sector, χ_t represents the contemporary productivity shock in the financial sector, while ν_{t+1} and μ_{t+1} are Gaussian white noises with constant standard deviations. T_t is a lump-sum transfer from taxation on the financial intermediary's profit, which mimics the

[18] I use non-capital letters for the individual household's optimization variables, and capital letters for the aggregate variables. To distinguish from their counterparts in the financial intermediary's profit maximization, I add a superscript "h" to both the net purchase of firm's shares s_t^h and the supply of resources in the intratemporal borrowing and lending market x_t^h.

results of a Nash bargaining between the unity of households and the financial intermediary.

The optimization states that at each period t, after the realization of the productivity shocks, an individual household takes the contemporary interest rate of capital r_t, the profit of intermediate goods firms π_t, the price of newly created or survived incumbent intermediate goods firms Q_t, and the return rate of lending to the financial intermediary R_t as given, and maximizes its life-long expected utility, subject to the budget constraint (1.13), an resource constraint (1.14), the evolution of aggregate productivity shocks (1.15), and the non-negativity constraint (1.16). As usual, net investment into the capital stock is defined by

$$i_t = k_{t+1} - (1-\delta)k_t. \tag{1.17}$$

Net purchase of firm's shares is defined by

$$s_t^h = a_{t+1} - (1-\eta)a_t. \tag{1.18}$$

In order to understand the necessity of the resource constraint (1.14), remember that the financial market for borrowing and lending between the financial intermediary and the households operates intratemporarily. That implies each household has a perfectly elastic supply function of resources. It is necessary for me to put a natural upper bound for how much an individual household could afford to lend, which in this case is the total wealth a household owns. In equilibrium, though, both the total and per-capita size of borrowing and lending shall be determined by the demand side as in equation (1.8) and (1.9), and (1.14) will never be binding.

First order conditions of the individual household's optimization give the following three equations

$$\mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)}(1+r_{t+1})\right] = 1, \tag{1.19}$$

$$Q_t = \mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)}\Big(\pi_{t+1} + (1-\eta)Q_{t+1}\Big)\right], \tag{1.20}$$

$$R_t = 1. \tag{1.21}$$

They are the non-arbitrage conditions for investment into the capital stock, investment into the shares of firms, and lending to the financial intermediary respectively. Equation (1.20) provides the foundation for the pricing equation (1.4). To see it is the non-arbitrage condition for firm's shares, notice what's on the left hand side is the cost of purchasing one unit share of firms today, and what's on the right is the expected return of owning one unit share of firms tomorrow, they must be equalized in equilibrium.

1.2.6 Aggregation and the Equilibrium

The aggregation of the economy is straightforward.[19] Since the measure of households is normalized to 1, I have

$$C_t = c_t,\ I_t = i_t,\ S_t^h = s_t^h,\ X_t^h = x_t^h,\ K_t = k_t,\ A_t = a_t.$$

I close the economy with the final goods market clearing condition

$$C_t + I_t + X_t^* + M_t^* = Y_t, \tag{1.22}$$

the capital market clearing condition

$$\int_0^{A_t} \Big(\exp(-\varepsilon_t) y_t^*(\omega)\Big)^{\frac{1}{\alpha}} d\omega = K_t, \tag{1.23}$$

[19] I use capital letters to denote the aggregate variables and non-capital letters to denote the per-capita variables.

the borrowing and lending market clearing condition

$$X_t^* = X_t^h, \tag{1.24}$$

and the firm's shares market clearing condition

$$S_t^h = S_t^*. \tag{1.25}$$

I give the following definition of equilibrium of the economy, see a full characterization of it in Appendix 1.6.6.

Definition. *A* ***D****ynamic* ***S****tochastic* ***G****eneral* ***E****quilibrium of the economy is a set of state-dependent prices, aggregate variables, and individual variables*

$$\Bigg\{ \underbrace{r_t, p_t, Q_t, R_t}_{prices} \; ; \; \underbrace{K_t, A_t, \varepsilon_t, \chi_t}_{state\ variables} \; ; \; \underbrace{c_t, k_{t+1}, a_{t+1}, x_t^h}_{households} \; ; \; \underbrace{X_t^*, e_t^*, M_t^*, S_t^*}_{financial\ intermediaries} \; ; \underbrace{Y_t, y_t, \pi_t}_{firms} \Bigg\}_{t=0}^{\infty}$$

such that

Given r_t and the demand function, $\{y_t, \pi_t\}$ solve the intermediate goods firm's problem;

Given Q_t and R_t, $\{X_t^*, e_t^*, M_t^*, S_t^*\}$ solve the individual financial intermediary's problem;

Given r_t, Q_t and R_t, $\{c_t, k_{t+1}, a_{t+1}, x_t^h\}$ solve the individual household's problem.

The rest of the variables are pinned down by corresponding equations. Aggregate state variables evolve according to their law of motions. All the aggregation and market clearing conditions are met.

1.3 Data and Empirical Evidence

I choose the U.S. initial public offering (IPO) market as a representative market for testing my theory. The IPO market fits my model of financial intermediation well for several reasons.

Firstly, an IPO is typically when the former private shareholders of the firm are getting paid, and it is a necessary prerequisite for the firm's shares to be traded publicly among households. Secondly, unlike other real-life intermediation activities that may take years to process and mature, the IPO is usually more time-sensitive and prompt to changes in the economic conditions. Thirdly, the commission fee of the investment bankers helping issue IPO is affected by the expectation of the quality of the issuing firm, as well as the long-term performance of its stock. So as in the model, the commission rate of the investment bankers should contain information about their effort intensity, as well as the quality of services they provide.

One more thing to clarify before going to details of the data. I call the IPO market a "representative market" in the sense that those investment bankers are part of the financial intermediary I have in the model. Thus the commission fee they get should also be a portion of what's in the model. I'm assuming that the portion is fixed, so I can still calculate the correlations of variables precisely, but not the levels. This is why in the quantitative part, I'm only going to use the correlations of variables found in the data, instead of their levels.

The data source for the U.S. IPO market is the SDC: Global New Issues Database. For what I need, it provides: issue date, gross spread amount,[20] principal amount,[21] offer price, type of the security, etc. I'm only looking at IPOs in the U.S. market, issued by U.S. firms and issuing common stocks,[22] between 01/01/1976 and 12/31/2016. There are 11509 stocks with all the information needed available. I choose the U.S. market because related studies show that investment bankers in the U.S. have been frequently colluding since the 70s, to avoid price competition. They as a group has certain monopoly power, so I don't need to worry about large changes in market power over time.

To document their correlations with the economic fundamentals, I construct quarterly

[20]The total amount of money paid to the investment bankers.
[21]One commonly accepted measure of the total value of an IPO.
[22]This includes common stock, ordinary stock, class A stock, and class B stock.

time series of the aggregate value, number, and commission rate[23] of IPOs. For each quarter in 1976Q1-2016Q4, I define the *quarterly total value* of IPOs as

$$\text{value}_t = \sum_i \text{Principal Amount}_t^i \times \text{Inflation Adjustment}_t,$$

which maps to $Q_t \cdot S_t^*$. The principal amount measures the value of a firm on its offering price, that is, the value in the primary market. I regard it as a fair price paid to obtain the ownership of the firm. The inflation adjustment is constructed using CPI (1=2010$). As in the model, I denote the *quarterly total number* of IPOs as S_t^*.

I define the *quarterly average commission rate* as

$$\text{acmr}_t = \frac{\sum_i \text{Gross Spread Amount}_t^i}{\sum_i \text{Principal Amount}_t^i},$$

which is a measure of the average commission fee paid to the investment bankers per unit value of IPO issued within the quarter.

I construct two variables as economic fundamentals. The first one is the cyclical component of real GDP. I take an HP filter with smoothing parameter 1600 on the log of quarterly real GDP to get $\ln y_t^c$. I define the GDP's deviation from trend as $\text{dev}_t = 1 - \exp(\ln y_t^c)$. The second one is the unemployment rate. I construct a quarterly series of unemployment rate ur_t by taking average of the BLS monthly data.[24]

From the table 1.1, one can see that the quarterly total value of IPO is procyclical, while the quarterly average commission rate is countercyclical, consistent with the model's prediction. The quarterly total number of IPO is acyclical, this is because the quarterly average value of IPO is highly procyclical.

[23] Speaking from an accounting point of view, the total payment that investment bankers receive from IPO includes not only just commission fee. But for simplicity, I call all those payments commission fee in this paper.

[24] Civilian unemployment rate (U-3), seasonally adjusted.

Table 1.1: Correlations between Variables of Interest

Correlation	Cyclical GDP ($\ln y_t^c$)	Unemployment Rate (ur_t)
Commission Rate (acmr_t)	-0.21	0.24
Total Value (value_t)	0.15	-0.01
Total Number (S_t^*)	0.06	0.04

The IPO literature states that an individual IPO's commission rate is negatively correlated with its value. For example, in Chen and Ritter (2000), Ritter (2003), they ran regressions of the individual IPO's commission rate on value, as well as other firm-specific characteristics. The coefficient of the individual value term is always negative and significant. To control for the individual IPO value effect, I define the *quarterly average value* of IPO as

$$\mathrm{avalue}_t = \frac{\mathrm{value}_t}{S_t^*}.$$

This is a measure of the value of a representative IPO issued within the quarter, and it maps to Q_t in the model.

I further divide those 11509 stocks into two subgroups: information-creating (I) group and non-information-creating (NI) group. The division is based on three key factors: 1. The offer price; 2. The primary market where the stock is listed or traded; 3. The firm's main business. The idea behind the division is to get rid of the "penny stocks", which are not traded publicly even after their IPOs. I put them in the NI group and use it as a control group. Detailed selection criteria can be found in Appendix 1.6.7, and I end up with 7605 information-creating stocks and 3904 non-information-creating stocks.

I run OLS regressions of the quarterly average commission rate on the quarterly average value, and each of the fundamentals, as shown in table 1.2.

Table 1.2: Regression Results on the Quarterly Average Level

Dep. Var.: acmr, **information-creating group**, 1976Q1-2016Q4[25]

	(1)	(2)	(3)
avalue	-3.77^{***}	-3.77^{***}	-3.61^{***}
$\ln y^c$	-6.81^{*}		
dev		-6.76^{*}	
ur			0.08^{**}
R^2	0.54	0.54	0.55
NO.	163	163	163

$[p < 0.05(*),\ p < 0.01(**),\ p < 0.001(***)]$

Dep. Var.: acmr, **non-information-creating group**, 1976Q1-2016Q4[26]

	(1)	(2)	(3)
avalue	-4.62^{***}	-4.62^{***}	-4.71^{***}
$\ln y^c$	9.07		
dev		9.32	
ur			0.09
R^2	0.46	0.46	0.46
NO.	163	163	163

$[p < 0.05(*),\ p < 0.01(**),\ p < 0.001(***)]$

The regression results show that the average commission rate in the information-creating group is countercyclical even after controlling for the average value of IPO. While its counterpart in the non-information-creating group doesn't show such a pattern. As a robustness check, I've also run regressions with individual stocks rather than the quarterly aggregates, the cyclical patterns stay the same. Regression results regarding individual stocks are provided in Appendix 1.6.8.

To summarize, I have documented the fact that in the U.S. IPO market, the total value of IPO is procyclical, while the commission rate of investment bankers is countercyclical, even after controlling for the average or individual IPO value. These empirical results, especially the correlations are consistent with my model's prediction, and will be used in the following quantitative part of the paper.

1.4 Quantitative Results

In Section 1.2, I lay out the model and provide several important qualitative results regarding the correlation between the optimal effort intensity (effort-to-size ratio) and the investment size. Most importantly, I have established the existence of an "excess region" of the investment size. In that region, the effort intensity, the average quality of financial services (success-to-size ratio), and the commission rate (commission-to-size ratio) are all negatively correlated with the investment size. The negative correlation implies a countercyclical efficiency of the financial intermediary in utilizing resources, which lays the foundation of the dampening effect that I'm about to examine quantitatively in the current section.

To take the theory to the data, I calibrate the model to the U.S. economy, including the correlation between the cyclical GDP and the investment size documented in Section 1.3. I want to check how much the calibrated model can explain the documented correlation between the cyclical GDP and the commission rate. I also conduct several counterfactual exercises, to show that the financial intermediary's cyclical behavior in exerting screening effort helps reduce the output and household consumption volatility, thus a dampening effect. Let's start with a specification of functional forms for the remaining functions.

1.4.1 Model Specification

I have specified the functional forms for the cost of screening effort $F(e)$ and the technology of handling resources $n(X)$ in Assumption 4. To proceed, I need to specify the population distribution of abilities $H(z)$, and the household's utility function $u(c)$.

Assumption 5. *(Functional Form Specification)*

$$H(z) = z,\ \forall z \in [0,1].$$

$$u(c) = \frac{c^{1-\psi}}{1-\psi};\ \psi > 0.$$

That is, I assume a uniform distribution of abilities and a constant relative risk aversion utility function. The distribution function is consistent with Assumption 1. I use these specifications because they allow for closed-form solutions. The main qualitative results would remain with other specifications.

Since $H(z) = z$, one can easily derive that $\mathbb{E}(z) = \frac{1}{2}$, $\Delta_{H(z)} = \frac{1}{8}$, $\underline{z}(N,e) = 1 - \frac{N}{e}$, and the effort cutoff is $2N$. I give the maximum success function as

$$S(N,e) = \begin{cases} \frac{e}{8} + \frac{N}{2}, & \frac{e}{N} < 2 \\ N\Big(1 - \frac{N}{2e}\Big), & \frac{e}{N} \geq 2 \end{cases}. \tag{1.7'}$$

The marginal benefit of effort is

$$Q \bullet S_e(N,e) = \begin{cases} \frac{Q}{8}, & \frac{e}{N} < 2 \\ \frac{Q}{2}\Big(\frac{N}{e}\Big)^2, & \frac{e}{N} \geq 2 \end{cases}.$$

One can check that these two equations are consistent with what has been described in Proposition 1 and Corollary 2.

Using the first order condition $Q \bullet S_e(N,e) = F'(e)$, I can solve for the effort policy function in the two-step maximization problem

$$e(N) = \begin{cases} \left(\dfrac{Q \bullet N^2}{2f}\right)^{\frac{1}{2+\kappa}}, & N \leq \dfrac{1}{2}\left(\dfrac{Q}{8f}\right)^{\frac{1}{\kappa}} \\ \left(\dfrac{Q}{8f}\right)^{\frac{1}{\kappa}}, & N > \dfrac{1}{2}\left(\dfrac{Q}{8f}\right)^{\frac{1}{\kappa}} \end{cases}. \tag{1.26}$$

The threshold value of N is $\bar{N}(Q) = \dfrac{1}{2}\left(\dfrac{Q}{8f}\right)^{\frac{1}{\kappa}}$. One can check that $e(N)$ is consistent with what has been described in Theorem 3. The corresponding threshold value of X is $\bar{X}(Q) = \left[\exp(-\chi)\xi \bullet \bar{N}(Q)\right]^{\frac{1}{\lambda}}$. Different from $\bar{N}(Q)$ though, $\bar{X}(Q)$ is increasing in Q but decreasing in χ, so the cyclicality of $\bar{X}(Q)$ is ambiguous.

Again, suppose the equilibrium X^* and e^* are such that $\dfrac{e^*}{N^*} < 2$, I have

$$X^* = \left(\frac{\exp(\chi)\lambda Q}{2R\xi}\right)^{\frac{1}{1-\lambda}},$$

$$e^* = \left(\frac{Q}{8f}\right)^{\frac{1}{\kappa}}.$$

And the condition under which $\dfrac{e^*}{N^*} < 2$ shall be met is

$$Q > \left[2^{\lambda\kappa-(1-\lambda)(3+\kappa)}\left(\frac{\xi}{\exp(\chi)}\right)^{\kappa}\left(\frac{1}{f}\right)^{1-\lambda}\left(\frac{R}{\lambda}\right)^{\lambda\kappa}\right]^{\frac{1}{\lambda\kappa-(1-\lambda)}}.$$

The equilibrium X^* and e^* in the other scenario, namely when $\dfrac{e^*}{N^*} > 2$, can be solved numerically from the first order conditions (1.8) (1.9). Unfortunately, they don't have a closed-form solution like the ones here.

Table 1.3: Benchmark Calibration

	Param	Value	Target
Firms	α	0.38	capital share
	δ	0.012	Cooley and Prescott (1995)
	η	0.02	BDS statistics
	σ	6	average markup
	τ	0	equilibrium stability
Households	β	0.987	Cooley and Prescott (1995)
	ψ	2	literature
Shocks	std(ν)	1.3×10^{-5}	std GDP
	ρ_{11}	0.71	autocorr GDP
	std(μ)	2.7×10^{-5}	corr GDP & IPO size
	ρ_{22}	0.68	autocorr IPO size
	ρ_{12}, ρ_{21}	0	no structural corr
Intermediary	γ	0.03	IPO commission rate
	ξ	9.72	S&P 500 P/B ratio
	κ	0.59	higher moments
	λ	0.82	profit rate
	f	10.30	higher moments
	f^0	0	set

1.4.2 Calibration

I calibrate the model to the U.S. economy, as well as some important correlations found in the empirical part of the paper. I have in total of 19 parameters to calibrate. Among those some I calibrate to match the long-term mean of the economy, some I calibrate jointly using the simulated method of moments. To be consistent with the real business cycle literature, parameters are calibrated on the quarterly frequency. Here is a table for my current benchmark calibration.

For the firm's part, I calibrate the concavity parameter α to match the long-term average capital share calculated for the U.S. economy. I take the capital depreciate rate δ from Cooley and Prescott (1995). I calibrate the death rate of incumbent intermediate goods firms η to match the quarterly average death rate of firms reported by the BDS statistics. I calibrate the elasticity of substitution between intermediate goods σ to match the long-term average markup, calculated for the U.S. economy in another working paper of mine: "Antitrust Policy in a Globalized Economy". At last, I set the *love of variety* parameter τ to be 0 because the model won't have a stable equilibrium with a large τ.

For the household's part, I take the time discount factor β from Cooley and Prescott (1995), and the risk aversion parameter ψ from the literature.

For the productivity shock in the intermediate goods sector, I calibrate the standard deviation of the white noise ν and the self-persistence parameter ρ_{11} jointly, using the simulated method of moments, to match the standard deviation and autocorrelation of the cyclical real GDP. For the productivity shock in the financial sector, I calibrate the standard deviation of the white noise μ and the self-persistence parameter ρ_{22} jointly, using the simulated method of moments, to match the correlation between the cyclical real GDP and the quarterly total size of IPO found in table 1.1, and the autocorrelation of the quarterly total size of IPO.[27] I set the cross-persistence parameters $\rho_{12} = \rho_{21} = 0$ because I want to rule out structural correlations between productivity shocks in different sectors.

For the financial intermediary's part, I calibrate the taxation (bargaining weight) parameter γ to match the 7% average commission rate in the U.S. IPO market. I calibrate the cost of industrialization ξ to match the average S&P 500 "Price-to-Book" ratio.[28] The calibration for the concavity parameter λ in the technology of handling resources, and parameters governing the cost of effort κ and f, is still work in progress. I set the fixed cost of effort

[27]Notice that when it comes to the statistics of the IPO market, I use only normalized terms: correlation, autocorrelation, instead of levels such as standard deviation. I have explained the reason in the empirical part of the paper.

[28]In the model, $\frac{Q_t}{\xi}$ can be regarded as the "Price-to-Book" ratio.

$f^0 = 0$ and it is a trivial parameter in my model.

Under the benchmark calibration, I have solved the model numerically, and simulated time series of cyclical GDP Y, investment size X^*, effort level e^*, commission fee M^*, etc. I'm particularly interested in how well the calibrated model can explain the documented correlation between the cyclical real GDP and the commission rate.

Table 1.4: Simulated Correlations under the Benchmark

Correlation with Y	M^*/X^*	e^*/X^*	S^*/X^*	C
Simulation	-0.16	-0.18	-0.33	0.74
Data	-0.21	n/a	n/a	0.76

The simulated correlation between the cyclical real GDP and the commission rate explains 71% of the correlation documented in the data. The model also provides simulated correlations regarding the effort intensity and the financial intermediary's efficiency in utilizing resources, which are difficult to observe directly from the data.[29]

At last, to examine whether the financial intermediary's cyclical behavior in exerting screening effort helps dampen the economic volatility, I conduct several counterfactual exercises.

Table 1.5: Counterfactual Results under the Benchmark

Counterfactual		Y Volatility	C Volatility	Household Welfare
Vary $F(e)$	20%	0.14%	0.32%	-2.22%
	-20%	-0.24%	-0.53%	3.84%
Commission Rate Cap		0.36%	0.54%	-5.62%
Shut Down Intermediary		0.74%	1.55%	-7.03%

[29]There is hope in measuring S^* by the long-term performance of the firms after IPO. I'm working on it as an extension of the empirical exercises of the paper.

As shown in table 1.5, in the first counterfactual exercise, I raise (reduce) the financial intermediary's cost of effort by 20%. The output volatility[30] increases (decreases) by 0.14% (0.24%). The household consumption volatility increases (decreases) by 0.32% (0.53%). In addition, the households' welfare loss (gain) is 2.22% (3.84%).[31]

As for the second exercise, I put a binding[32] commission rate cap on the financial intermediary. The output volatility increases by 0.36%. The household consumption volatility increases by 0.54%. And the households' welfare loss is 5.62%. Based on this counterfactual exercise, I argue that it is inefficient to drive down the commission rate level by using a price cap. A better way is to introduce competition to the financial market, which is equivalent to reducing γ in the model.

Lastly, I shut down the financial intermediary completely by setting $e \equiv 0$ and $\gamma = 0$. The output volatility increases by 0.74%. The household consumption volatility increases by 1.55%. And the households' welfare loss is 7.03%. From the counterfactual exercises, I conclude that the cyclical patterns of the financial intermediary's behaviors, especially its effort intensity in screening entrepreneurs, help dampen the volatility of output and household consumption, and improve welfare.

The values of λ, κ and f are not calibrated yet, and I simply pick ad hoc numbers to put in the benchmark calibration table. From the closed-form solutions of X^* and e^*, one can make the conjecture that κ shall largely affect the dampening effect found in the previous counterfactual exercises. Eventually, I will calibrate those parameters to the data. But for now, let me do a robustness check with a much higher $\kappa = 3$. The corresponding simulation and counterfactual results can be found in Appendix 1.6.9.

[30]Measured by the standard deviation to mean ratio, aka the coefficient of variation, because the variable's mean changes a lot too.

[31]The welfare changes are mainly (around 90%) due to changes in the steady-state level of consumption.

[32]The cap is set to be 3%, which matches the commission rate level in the European IPO market, and is much lower than the model's steady-state commission rate level.

1.5 Conclusion

Financial intermediation helps direct funds to its best uses. Financial intermediaries' reactions to changes in the economic fundamentals over business cycles may in turn reshape the cycles. What are the cyclical patterns of the financial intermediaries' behaviors? Such as their investment size, effort in screening entrepreneurs (ideas), efficiency in utilizing resources, etc. To answer this question, I develop a theory of financial intermediation under a stochastic general equilibrium framework. Given the cyclical patterns of the financial intermediaries' behaviors, how would they reshape business cycles? To answer the second question, I take the model to the data and conduct quantitative analysis. Let me summarize the main results here.

The paper provides new findings and perspectives in theory, empirical evidence, and quantitative analysis. In terms of theory, I build an innovative model of financial intermediation over business cycles to show that there could exist a dampening effect of the financial intermediary. The dampening effect is originated from the financial intermediary's optimal behaviors over the cycles. Under certain conditions, the financial intermediary's effort intensity in screening entrepreneurs is negatively correlated with its investment size. This results in a higher (lower) efficiency in utilizing resources in bad (good) times. It also implies that the commission rate of the financial intermediary should be higher (lower) in booms (recessions). I'm able to deliver these qualitative results under a fairly general setup, which can be applied to many more specialized applications.

In the empirical part, I have examined one application of the model: the cyclicality of the commission rate. I have documented the fact that the commission rate for investment bankers in the U.S. IPO market is countercyclical, which is consistent with my model's prediction. This indicates that, on average, they put more effort into screening startup firms and get paid more per unit of investment in recessions, which helps the economy to recover.

As for the quantitative analysis, after calibrating the model to the U.S. economy, I have shown that The cyclical behaviors of the financial intermediary help dampen the volatility of output, household consumption, and improve household welfare. A 20% drop in the financial intermediary's cost of effort reduces the output volatility and the household consumption volatility by 0.24%, 0.53% respectively. The households' welfare gain is 3.84%. I have also shown that price-cap is an inefficient way of reducing the level of commission fee, a better way could be introducing more competition to the financial market.

A natural extension of the model is to introduce a more active entrepreneur part. For example, I'm trying to add a labor market into the model, and the workers choose between working in an intermediate goods firm or becoming an entrepreneur to produce ideas. The main difficulty is tractability since it raises a nontrivial principal-agent problem between the financial intermediary and the entrepreneurs. But this is surely an interesting direction to explore on.

1.6 Appendix

1.6.1 Deriving the Success Function

Following **Definition 1**, for a pair of (N, e), there could be two scenarios. First is that $e\big[1 - H\big(\mathbb{E}(z)\big)\big] < N$, in this scenario the pool of verified entrepreneurs is not enough to absorb all the borrowed resources. Thus $e\big[1 - H\big(\mathbb{E}(z)\big)\big]$ many verified entrepreneurs will get funded, and the rest are randomly drawn, unverified entrepreneurs. This result in an aggregate number of success as

$$\begin{aligned} S(N, e) &= \underbrace{e\int_{\mathbb{E}(z)}^{1} zh(z)dz}_{\text{verified}} + \underbrace{\mathbb{E}(z)\Big(N - e\big[1 - H\big(\mathbb{E}(z)\big)\big]\Big)}_{\text{unverified}} \\ &= e\int_{\mathbb{E}(z)}^{1} \big(z - \mathbb{E}(z)\big)h(z)dz + N \bullet \mathbb{E}(z). \end{aligned}$$

The second scenario is that $e\big[1 - H\big(\mathbb{E}(z)\big)\big] \geq N$, where all resources go to the verified entrepreneurs in a descending order. And the threshold for getting funded is an $\underline{z}$ solved from

$$e\big[1 - H(\underline{z})\big] = N.$$

One can easily derive that $\underline{z}(N, e) = H^{-1}\left(1 - \dfrac{N}{e}\right)$, which is higher than $\mathbb{E}(z)$ in this scenario, and the resulted aggregate number of success is

$$S(N, e) = e\int_{\underline{z}(N,e)}^{1} zh(z)dz.$$

1.6.2 Proof of Proposition 1 and Corollary 2

Let's start by deriving the first order partial derivatives of $S(N,e)$. Under Assumption 1, I know that $\underline{z}(N,e) = H^{-1}\Big(1 - \dfrac{N}{e}\Big)$ is differentiable. One can derive

$$\frac{\partial \underline{z}(N,e)}{\partial N} = -\frac{1}{eh(\underline{z})},$$

$$\frac{\partial \underline{z}(N,e)}{\partial e} = \frac{N}{e^2 h(\underline{z})}.$$

For notational purpose, I define

$$E_L = \left\{ (N,e) \in \mathbb{R}^2_+ \,\middle|\, \frac{e}{N} < \frac{1}{1 - H\big(\mathbb{E}(z)\big)} \right\}$$

and

$$E_H = \left\{ (N,e) \in \mathbb{R}^2_+ \,\middle|\, \frac{e}{N} > \frac{1}{1 - H\big(\mathbb{E}(z)\big)} \right\}.$$

Obviously $S(N,e)$ is continuously differentiable on E_L and E_H. I can derive

$$S_N(N,e) = \begin{cases} \mathbb{E}(z), & \dfrac{e}{N} < \dfrac{1}{1 - H\big(\mathbb{E}(z)\big)} \\[2ex] \underline{z}(N,e), & \dfrac{e}{N} > \dfrac{1}{1 - H\big(\mathbb{E}(z)\big)} \end{cases} \tag{1.27}$$

and

$$S_e(N,e) = \begin{cases} \displaystyle\int_{\mathbb{E}(z)}^{1} \big(z - \mathbb{E}(z)\big) h(z) dz, & \dfrac{e}{N} < \dfrac{1}{1 - H\big(\mathbb{E}(z)\big)} \\[2ex] \displaystyle\int_{\underline{z}(N,e)}^{1} \big(z - \underline{z}(N,e)\big) h(z) dz, & \dfrac{e}{N} > \dfrac{1}{1 - H\big(\mathbb{E}(z)\big)} \end{cases}. \tag{1.28}$$

When $e \xrightarrow{+} \frac{N}{1-H\left(\mathbb{E}(z)\right)}$, $\underline{z}(N,e) \xrightarrow{+} \mathbb{E}(z)$. This indicates that both $S_N(N,e)$ and $S_e(N,e)$ are continuous on $\mathbb{R}^2_+$, so $S(N,e)$ is continuously differentiable on $\mathbb{R}^2_+$. In addition, since $S_N(N,e) > 0$ and $S_e(N,e) > 0$, $S(N,e)$ is strictly increasing in both N and e on $\mathbb{R}^2_+$.

Next is to examine the curvature of $S(N,e)$. Let's take second order partial derivatives

$$S_{NN}(N,e) = \begin{cases} 0, & \dfrac{e}{N} < \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \\ -\dfrac{1}{eh(\underline{z})}, & \dfrac{e}{N} > \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \end{cases},$$

$$S_{Ne}(N,e) = \begin{cases} 0, & \dfrac{e}{N} < \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \\ \dfrac{N}{e^2 h(\underline{z})}, & \dfrac{e}{N} > \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \end{cases},$$

$$S_{ee}(N,e) = \begin{cases} 0, & \dfrac{e}{N} < \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \\ -\dfrac{N^2}{e^3}\dfrac{1}{h(\underline{z})}, & \dfrac{e}{N} > \dfrac{1}{1-H\big(\mathbb{E}(z)\big)} \end{cases}.$$

From those partial derivatives one can conclude that $S(N,e)$ is linear on E_L. To see that it is strictly concave on E_H, for an arbitrary vector $(x_1,x_2) \in \mathbb{R}^2$,

$$\begin{aligned} & S_{NN}x_1^2 + 2S_{Ne}x_1x_2 + S_{ee}x_2^2 \\ = & -\frac{1}{eh(\underline{z})}\Big(x_1^2 - \frac{2N}{e}x_1x_2 + \frac{N^2}{e^2}x_2^2\Big) \\ = & -\frac{1}{eh(\underline{z})}\Big(x_1 - \frac{N}{e}x_2\Big)^2 \leq 0 \end{aligned}$$

That is, the Hessian matrix of $S(N,e)$ is negative definite except for a measure 0 of points, which proves its strict concavity. Proposition 1 proved.

The last is to see how changes of N affect $S_e(N,e)$. Notice that on E_L, $S_e(N,e)$ is a constant w.r.t. both N and e, thus $S(N,e)$ is linear in e with a slope that is constant w.r.t. changes of N.

However on E_H, $S_e(N,e)$ is strictly decreasing in both e and N. To summarize, an increase of N enlarges the constant region of $S_e(N,e)$, and shifts $S_e(N,e)$ up on its decreasing region. Loosely speaking, it shifts $S_e(N,e)$ to the right.

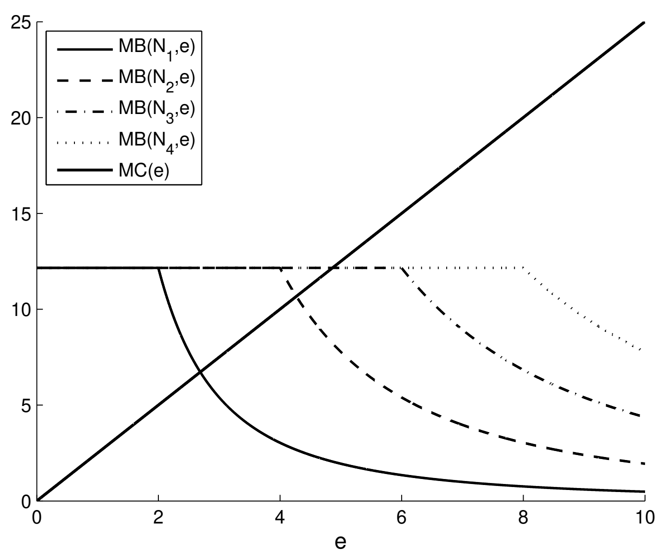
Once I take into consideration of the firm's price Q, it shifts up the marginal benefit of effort $Q \cdot S_e(N, e)$ proportionally. Corollary 2 proved.

1.6.3 Proof of Theorem 3

Using results from Proposition 1 and Corollary 2, the proof is straightforward. Let's start with the existence and uniqueness of $e(N)$. As shown in the proof of Proposition 1, $S_e(N, e) > 0$. Assumption 2 states that $F'(0) = 0$ and the fixed cost component is very small, so the optimal effort level $e(N)$ must be an interior solution. In addition, since $S(N, e)$ is concave while $F(e)$ is strictly convex, $e(N)$ would be uniquely pinned down by the first order condition

$$\underbrace{Q \cdot S_e(N, e)}_{\text{marginal benefit}} = \underbrace{F'(e)}_{\text{marginal cost}} .$$

Figure 1.4: Marginal Benefit v.s. Marginal Cost



Depending on different values of N, there could be two scenarios. First is that N is small so $F'(e)$ intersects with $Q \bullet S_e(N, e)$ above the effort cutoff $\frac{N}{1-H\left(\mathbb{E}(z)\right)}$. For example, N_1 or N_2 in the graph above. If that is the case, an increase of N will raise e up. Because as shown in Corollary 2, an increase of N shifts $Q \bullet S_e(N, e)$ up when e is above the effort cutoff.

The second scenario is that N is large enough so $F'(e)$ intersects with $Q \bullet S_e(N, e)$ below the effort cutoff. In this case, when N increases, the intersection point will not be affected. Because an increase of N has no effect on $Q \bullet S_e(N, e)$ when e is below the effort cutoff.

Also notice that an increase of N raises the effort cutoff, so it is basically shifting $Q \bullet S_e(N.e)$ to the right. There must exist a $\bar{N}$ such that once $N > \bar{N}$, the intersection is below the cutoff.

At last, since an increase of Q shifts $Q \bullet S_e(N, e)$ up proportionally everywhere, it raises e as well as the $\bar{N}$.

1.6.4 Centralized Financial Market

In the centralized setup, I assume a collective contract between the unity of households and the financial intermediary governing the borrowing size X and effort level e, and a Nash bargaining to determine the return rate R.

Assumption 6. *The unity of households and the financial intermediary write a collective contract governing X and e. For any given Q, X, e and realization of χ, the promised return rate R, as well as the commission fee M, are determined by a Nash bargaining over the net surplus*[33]

$$Q \bullet S(N, e) - F(e) - X, \tag{1.5'''}$$

subject to $N = exp(\chi)\dfrac{n(X)}{\xi}$. The bargaining power of the financial intermediary is $\gamma \in (0, 1)$.

The surplus term is essentially the profit function of the financial intermediary in the decentralized financial market, since in the decentralized equilibrium the intratemporal re-

[33] At the optimal effort level it will always be non-negative.

turn rate to households R must be equalized to 1. γ works as the parameter governing the financial intermediary's market power. The lower γ is, the less market power the financial intermediary has. So the act of introducing competition to the market could be done by reducing γ.

One may concern whether the split of the surplus is implementable through an incentive compatible collective contract, since it is contingent on e but the effort level is hidden from the households. To answer this question, I give the following theorem.

Theorem 4. *Any Nash bargaining over the surplus can be implemented by an incentive compatible collective contract between the unity of households and the financial intermediary.*

The proof of Theorem 4 is given in Appendix 1.6.5. The intuition is that as long as the financial intermediary doesn't deviate from the descending allocation rule, the number of success S serves as a perfect signal of the effort level e. So a contract written on S is equivalent to a contract written on e.

With Nash bargaining, the incentives for the unity of households and the financial intermediary in choosing X and e, as well as the allocation rule, are perfectly aligned. That is, to maximize the surplus specified in (1.5'''). The unity of the households and the financial intermediary negotiate on the optimal investment size and the optimal screening effort level. Again, denote the optimal investment size as X^*, and the optimal effort level as e^*. Given all other things equal, the X^* and e^* here should be exactly the same as their counterparts in the decentralized equilibrium.

Now the commission fee is

$$M^* = \gamma\big[Q \bullet S(N^*, e^*) - X^*\big] + (1-\gamma)F(e^*), \tag{1.11'}$$

And the unity of households gets

$$Q \bullet S(N^*, e^*) - M^*,$$

which shall be distributed evenly among all households.

Different from what's in the decentralized setup though, once a collective contract is finalized, an individual household is no longer allowed to freely choose how much resources to supply to the financial intermediary. This is because Nash bargaining yields a positive intratemporal net return for the unity of households.

$$\begin{aligned} R_{\text{centralized}} - 1 &= \frac{Q \bullet S(N^*, e^*) - M^*}{X^*} - 1 \\ &= \frac{(1-\gamma)\big[Q \bullet S(N^*, e^*) - F(e^*) - X^*\big]}{X^*} > 0. \end{aligned}$$

Remember that the reason for introducing a centralized financial market is to be able to vary the market power of the financial intermediary. To see how it is done, one can derive the profit of the financial intermediary through Nash bargaining as

$$M^* - F(e^*) = \gamma\big[Q \bullet S(N^*, e^*) - F(e^*) - X^*\big].$$

Which is exactly γ portion of its counterpart in the decentralized equilibrium.

As a matter of fact, I can achieve the same result in the competitive financial market by posing a tax rate of $1-\gamma$ on the financial intermediary's profit. The taxation income shall be rewarded to the households as a lump-sum transfer

$$T = (1-\gamma)\big[Q \bullet S(N^*, e^*) - F(e^*) - R \bullet X^*\big],$$

which doesn't affect an individual household's incentive on how much to lend on the competitive intratemporal borrowing and lending market. One can check that

$$Q \bullet S(N^*, e^*) = M^* + R \bullet X^* + T$$

shall always hold.

1.6.5 Proof of Theorem 4

The key is to show that there exists a collective contract to incentivize the financial intermediaries to always follow the descending allocation rule. Because if they do, the resulted success function $S(N, e)$ serves as a perfect signal of e. Writing a contract in terms of S is thus equivalent to one written on e.

Consider the following simple contract written by the households to elicit a certain effort level $\hat{e}$. Given the aggregate size of entrepreneurs to be funded N, the households take the ownership of the firms, but promise to pay a financial intermediary certain amount of final goods M, which is larger or equal to the private cost $F(\hat{e})$ incurred, as long as the number of firms created by this financial intermediary is no less than $S(N, \hat{e})$, otherwise paying 0.[34]

With a simple contract described above, a financial intermediary's unique best response is to exert effort level $\hat{e}$ and follow the descending allocation rule. To see that, one should remember $\hat{e}$ is the minimum effort level required to create $S(N, e)$ many new firms. If a financial intermediary chooses an effort level low than $\hat{e}$, it will get paid 0 for sure. If the financial intermediary chooses an effort level higher than $\hat{e}$, it may get paid the same M but with a higher cost. And with the effort level being exactly $\hat{e}$, the financial intermediary can create $S(N, e)$ many firms only if it follows the optimal allocation rule, aka the descending allocation rule.

Now to fully implement the commission fee determined by the Nash bargaining, the

[34] I assume limited liability on the financial intermediary side.

households need to write a complex contract in which the payment M changes with e according to Assumption 6. Since the commission fee is strictly increasing in both the number of firms created S, and the inferred cost of effort $F(e)$, the financial intermediary's unique best response is still to stick with the descending allocation rule.

To summarize, the key features for the contract to work are: 1. Given N and e, $S(N, e)$ is achievable only through the descending allocation rule; 2. $S(N, e)$ is strictly increasing in e, thus serves as a perfect signal of the effort level; 3. The commission fee determined by the Nash bargaining is strictly increasing in S and $F(e)$.

1.6.6 Characterizing the Equilibrium

At each period t, after the realization of the productivity shocks ε_t and χ_t, the economy is characterized by four state variables: the aggregate capital stock K_t, the aggregate variety stock A_t, the realized productivity shock in the intermediate goods sector ε_t, and the realized productivity shock in the financial sector χ_t.

I claim that the dynamic stochastic general equilibrium of the economy is characterized by the following equations.

For $\forall t$ and $K_t, A_t, \varepsilon_t, \chi_t$,

$$Y_t = A_t^{\tau+1} y_t^*; \tag{1.29}$$

$$p_t = A_t^{\tau}; \tag{1.30}$$

$$y_t^* = \left(\frac{\sigma - 1}{\sigma} \frac{\exp(\varepsilon_t/\alpha)}{r_t + \delta} \alpha A_t^{\tau} \right)^{\frac{\alpha}{1-\alpha}}; \tag{1.31}$$

$$\pi_t = p_t y_t^* - (r_t + \delta)\big(\exp(-\varepsilon_t) y_t^*\big)^{\frac{1}{\alpha}}; \tag{1.32}$$

$$\mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)}(1 + r_{t+1})\right] = 1; \tag{1.33}$$

$$Q_t = \mathbb{E}_t\left[\frac{\beta u'(c_{t+1})}{u'(c_t)}\Big(\pi_{t+1} + (1-\eta) Q_{t+1}\Big)\right]; \tag{1.34}$$

$$R_t = 1; \tag{1.35}$$

$$Q_t \bullet S_e(N_t^*, e_t^*) = F'(e_t^*); \tag{1.36}$$

$$Q_t \bullet S_N(N_t^*, e_t^*) \bullet \exp(\chi_t)\frac{n'(X_t^*)}{\xi} = R_t; \tag{1.37}$$

$$N_t^* = \exp(\chi_t)\frac{n(X_t^*)}{\xi}; \tag{1.38}$$

$$S_t^* = S(N_t^*, e_t^*); \tag{1.39}$$

$$M_t^* = \gamma\big[Q_t \bullet S_t^* - R_t \bullet X_t^*\big] + (1-\gamma) F(e_t^*); \tag{1.40}$$

$$T_t = (1-\gamma)\big[Q_t \bullet S_t^* - F(e_t^*) - R_t \bullet X_t^*\big]; \tag{1.41}$$

$$C_t = c_t;\ K_t = k_t;\ A_t = a_t^h;\ X_t^h = x_t^h; \tag{1.42}$$

$$I_t = K_{t+1} - (1-\delta)K_t; \tag{1.43}$$

$$\begin{pmatrix} \varepsilon_{t+1} \\ \chi_{t+1} \end{pmatrix} = \begin{pmatrix} \rho_{11} & \rho_{12} \\ \rho_{21} & \rho_{22} \end{pmatrix} \begin{pmatrix} \varepsilon_t \\ \chi_t \end{pmatrix} + \begin{pmatrix} \nu_{t+1} \\ \mu_{t+1} \end{pmatrix}; \tag{1.44}$$

$$C_t + I_t + X_t^* + M_t^* = Y_t; \tag{1.45}$$

$$A_t \Big(\exp(-\varepsilon_t) y_t^* \Big)^{\frac{1}{\alpha}} = K_t; \tag{1.46}$$

$$X_t^* = X_t^h; \tag{1.47}$$

$$A_{t+1} - (1-\eta)A_t = S_t^*; \tag{1.48}$$

Equation (1.29) - (1.32) are from the final goods and intermediate goods firm's static profit maximization problem. Equation (1.33) - (1.35) are from the individual household's optimization problem. Equation (1.36) - (1.41) are from the financial intermediary's static profit maximization problem, with a taxation on profit. Equation (1.42) and (1.43) are the aggregation conditions. Equation (1.44) governs the evolution of the productivity shocks. Equation (1.45) - (1.48) are the market clearing conditions for final goods, capital, inratemporal borrowing and lending, and firm's shares market respectively.

The economy exhibits a saddle-path stable steady-state when the *love of variety* τ is close to 0. Following the real business cycle literature, I solve a dynamic stochastic general equilibrium around the steady-state, using perturbation method. I'm not going to write out the equations characterizing the steady-state, cause they are just repetitions of equation (1.29) - (1.48), without the productivity shocks of course.

1.6.7 Criteria for Group Division

I did the following selection, what's left are in the information-creating group, what's taken out is in the non-information-creating group.

First, I take out stocks whose offer price is below 5$. This is based on SEC's definition of a "penny stock". These stocks are not traded frequently or widely after the IPO, so they are still considered to be unknown or unavailable to the general public.

Second, I take out stocks that are not listed or traded on a "national securities exchange".[35] For the same reason that these stocks are not traded frequently or widely after IPO.

Third, I take out IPO that are issued by firms whose main business are: open/close-end funds, special financial vehicles, blank check companies (SPACs), REITs. Because these stocks are issued for special purposes.

1.6.8 Regression Results of Individual Stocks

I pool all individual stocks together, and run regression of their commission rates (cmr_t^i) on values (val_t^i) and the unemployment rate (ur_t).

[35]SEC has a detailed list of national securities exchanges.

Table 1.6: Regression of Individual Stocks: All Stocks

Dep. Var.: cmr, **all stocks**, 1976Q1-2016Q4

	(1)	(2)
val	-1.71^{***}	-1.68^{***}
ur		0.15^{***}
R^2	0.09	0.10
NO.	11509	11509

$[p < 0.05(*),\ p < 0.01(**),\ p < 0.001(***)]$

I do the same for the information-creating group.

Table 1.7: Regression of Individual Stocks: Information-creating Group

Dep. Var.: cmr, **information-creating group**, 1976Q1-2016Q4

	(1)	(2)
val	-1.73^{***}	-1.73^{***}
ur		0.03^{***}
R^2	0.13	0.13
NO.	7605	7605

$[p < 0.05(*),\ p < 0.01(**),\ p < 0.001(***)]$

And for the non-information-creating group.

Table 1.8: Regression of Individual Stocks: Non-information-creating Group

Dep. Var.: cmr, **non-information-creating group**, 1976Q1-2016Q4

	(1)	(2)
val	-1.62^{***}	-1.60^{***}
ur		0.23^{***}
R^2	0.08	0.09
NO.	3904	3904

$[p < 0.05(*),\ p < 0.01(**),\ p < 0.001(***)]$

1.6.9 Robustness Check

In this robustness check, I set the convexity parameter in the cost of effort to be $\kappa = 3$, which is about 5 times to its benchmark value 0.59. The corresponding table of correlations is as follows.

Table 1.9: Simulated Correlations under Large κ

Correlation with Y	M^*/X^*	e^*/X^*	S^*/X^*	C
Simulation	-0.18	-0.20	-0.26	0.74
Data	-0.21	n/a	n/a	0.76

The corresponding table of counterfactual results is as follows.

Table 1.10: Counterfactual Results under Large κ

Counterfactual		Y Volatility	C Volatility	Household Welfare
Vary $F(e)$	20%	0.02%	0.11%	-0.37%
	-20%	-0.03%	-0.15%	0.48%
Commission Rate Cap		0.43%	2.03%	-6.64%
Shut Down Intermediary		1.51%	4.01%	-3.79%

Comparing to the benchmark calibration, the wellness of fit of the correlation between Y and $\frac{M^*}{X^*}$ has improved. However, the magnitude of the dampening effect has been greatly weakened.

Chapter 2

Antitrust Policy in a Globalized Economy

Linyi Cao[1] Lijun Zhu[2]

2.1 Introduction

The measured markup, i.e. the difference between price and marginal cost, has been increasing since the 1980s in the US (De Loecker and Eeckhout (2017), Barkai (2016)). An increasing markup leads to efficiency loss as it lowers aggregate production and raises price comparing to the social optimal level. Further, if the rise of markup among firms varies, misallocation caused by markup differences among heterogeneous firms worsens. On the other hand, as an important tool to restrain monopoly power, antitrust policy in the U.S. has been relaxed, and number of Mergers and acquisitions has surged since the 1980s. This paper provides a framework to evaluate cost and benefits of antitrust policy in a global con-

[1]Department of Economics, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130 (email: *l.cao@wustl.edu*).

[2]Institute of New Structural Economics, Peking University (email: *lijunzhu@nsd.pku.edu.cn*).

text. M&A reallocate resources from small to large and typically more productivity firms, while also increase monopoly power of the latter. Optimal antitrust policy seeks a balance between the positive productivity effect and the negative markup effect.

We first measure markup among US public firms using the method proposed in De Loecker and Warzynski (2012) and De Loecker and Eeckhout (2017), and empirically establish a connection between M&A activity and markup at firm level. The average markup, weighted by cost of goods sold, increases from about 4% in 1980 to 16% in 2017. Comparing to 1980, the distribution of markup among public firms in 2015 clearly shifts to the right. By combining Compustat public firm data and SDC mergers and acquisition data, we find that increase of a firm's markup in the next year is about 2% higher, in relative terms, if this firm has merged or acquired at least one other firm in the current year. This amplification effect is stronger if the acquirer has a larger per-merger markup.

We then build a dynamic general equilibrium mode of M&A to analyze optimal antitrust policy. Our model builds on David (2017), which developed a model of Mergers and acquisitions in a firm dynamics framework. Firm heterogeneity is summarized by a productivity parameter, which determines firm size. The M&A market is characterized as a two-sided costly search and matching. To capture complementarity between acquirer and targets, M&A is assumed to increase the acquirer's post-merger productivity, while the magnitude of increase depends on the pre-merger size of the target. Upon a successful M&A, the target exits the economy.

The main difference of our model from David (2017) is that we incorporate heterogeneous markup into the framework. Monopoly power is an important dimension of mergers and acquisitions, and our model allows to study optimal antitrust policy. Firms with different size charge different markups. In the model, large firms face a less elastic demand and charge a higher markup. Mergers and acquisitions increase acquirers' size as well as post-merger markup. Heterogeneous markup across firms result in misallocation, and a larger dispersion due to increase in markup of large firms amplifies productivity loss due to misallocation.

Antitrust policy is modeled though a search cost function in the M&A market. That is, a stricter antitrust policy imposes a higher merger cost. Optimal antitrust policy seeks to balance the positive productivity and the negative markup effect. The productivity effect reallocates resource to larger and more productive firms and increases aggregate production, while the markup effect increases misallocation and lowers aggregate output. As both effects are reflected in the aggregate output, we use aggregate production/consumption to measure welfare. Initially strict, a more lenient antitrust policy might increase production as it allows more productivity-enhancing reallocation. This increase is eventually reversed as the M&A technology admits decreasing return to scale on the acquirer side. Therefore, after a certain point, a further relaxation of antitrust policy starts to decrease total output as the productivity effect becomes limited and the markup effect takes over.

We quantitatively show that total consumption (or/and output) is a hump-shaped function of the antitrust policy variable in the model. The optimal policy corresponds to the one that maximizes total consumption. In a work in progress, we are extending the model to an open economy, and aim to formalize the intuition that open to trade demands a more lenient optimal antitrust policy and explore its quantitative implications for aggregate markup and welfare.

The rest of the chapter is organized as the following: In section 2.2, we present empirical facts regarding markup and Mergers & Acquisitions, and show that there is positive effect of M&A on firms' markup. The model, first in a closed economy and then extended to an open economy framework, is shown in Section 2.3. Section 2.4 provides a quantitative analysis, while a concluding remark is offered in Section 2.5.

2.2 Facts

This section provides empirical facts regarding markup, firm size and M&As. We measure markup using the method proposed in De Loecker and Warzynski (2012) and De Loecker and Eeckhout (2017)[3]. The measurement bases on the following observation: firm's markup, i.e. ratio of price to marginal cost, is equal to the ratio of the output elasticity with respect to an input without significant adjustment cost, to the cost share of this input[4]. The latter can be directly calculated from firms balance sheet, while the former is estimated by using a control function approach from the production estimation literature (Olley and Pakes (1996), Levinsohn and Petrin (2003)).

2.2.1 The Rising Markups

We apply the method mentioned above and use Compustat data to estimate firm level markups[5]. There are in total 321,315 $firm \times year$ valid observations from 1951 to 2017[6]. After obtaining firm level markup, we average them across firms by using input cost as the weight. The resulted average markup is plotted in Figure 2.1[7]. With up and downs, the average markup in early 1980s is at about the same level as in 1950s. However, over the last 3-4 decades, the measured markup has greatly increased from about 4% in 1980 to about 16% in 2017.

[3]See appendix a brief introduction of the method

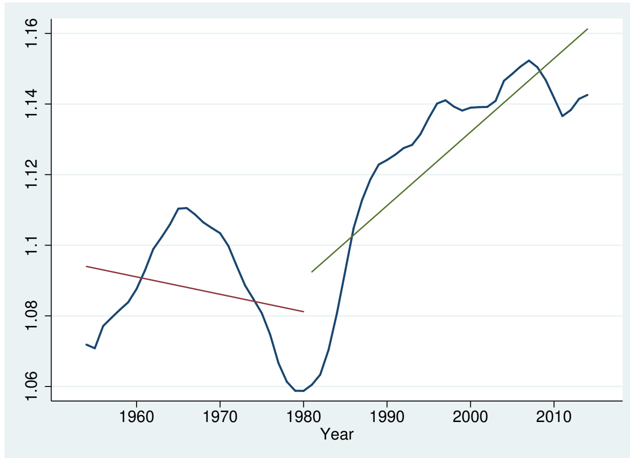
[4]This equivalence can be seen clearly from firm's cost minimization problem, as shown in appendix.

[5]Following De Loecker and Eeckhout (2017), we choose "cost of goods sold" in compustat as the choice input, and estimate output elasticity to this input for each 2 digit NAICS sector. In the baseline estimation, we assume this elasticity does not change over time.

[6]This number only includes $firm \times year$ observations for which data is available for the estimation of markup, and thus excludes observations with missing data. See Table 2.7 in appendix for NO in each year.

[7]Figure 2.5 in appendix shows both cost-weighted average markup and sales-weighted average markup.

Figure 2.1: Average Markups, 1950-2017



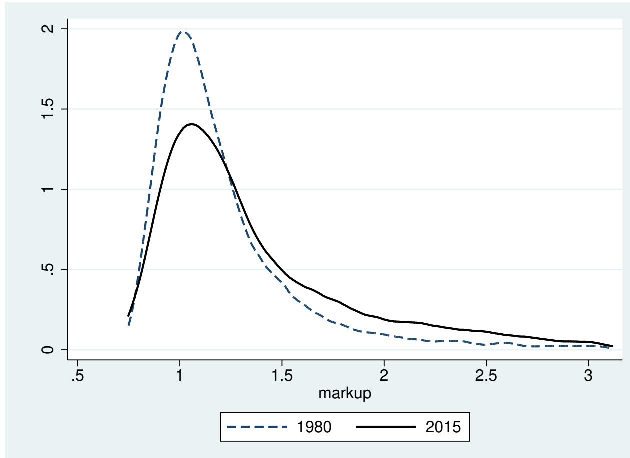
Note: Average markups among public firms, with weights equal to cost of goods sold. The value for year t is a moving average of original values from year $t-2$ to year $t+2$.*Data source:* Compustat.

Figure 2.2 presents the distribution of markup (i.e. estimated $\frac{\text{price}}{\text{marginal cost}}$) among Compustat public firms in 1980 and 2015[8]. The distribution in these two years show a clear difference[9]. The distribution in 2015 has more firms with markups on the right tail, while the 1980 distribution is characterized by a concentration of markups at the left tail. The median markup in 1980 is 1.10, 75% is 1.33 and 90% 1.76. These three numbers in 2015 are 1.20, 1.66, and 2.71 respectively. A rising average markup is mainly driven by a rise of markup on the right tail.

[8]We rank estimated markups among all public firms across all years in Compustat. To rule out extreme values, Figure 2.2 presents this distribution conditional on markup values locates between 5 and 95 percentiles. Distribution of markup values between 1 and 99 percentiles is shown in Figure 2.8 in Appendix.

[9]There are % of firms that have an estimated markup below 1, See De Loecker and for an explanation.

Figure 2.2: Distribution of Markups among Public Firms in 1980 and 2015



Note: Distribution of markups conditional on markup values between 5 and 95 percentiles of the whole distribution. *Data source:* Compustat.

Within a sector, large firms tend to have higher markups. Table 2.1 presents the correlation between firm size and measured markup. We simply run a regression of markup (in log) to a measurement of firm size. In particular, firm size can be measured using sale, capital, or employment. After controlling for year and sector fixed effect, the correlation between firms size and markup is positive and significant.

Table 2.1: Markup and Firm Size

Dependent Variable: $\ln(markup)$

	(1)	(2)	(3)
$\ln(sale)$	0.057***		
	(0.0005)		
$\ln(capital)$		0.025***	
		(0.0005)	
$\ln(emp)$			0.016***
			(0.0006)
Year D		Yes	
Sector D		Yes	
R^2	0.06	0.03	0.03
NO	321,339	321,339	286,658

Note: Data source: Compustat, 1980-2017

The fact that average markup has been rising is also related to increasing concentration in many industries in the U.S. since the 1980s, as documented in Autor, Dorn, Kats, Patterson, and Van Reenen (2017) and Boldrin and Zhu (2018). Figure 2.6 in appendix presents concentration index in the Manufacturing sector, which was relatively stable from the 1960s to the early 1980s, but has been increasing over the past three to four decades. The index for the aggregate economy, averaged over 2-digit NAICS sectors and shown in Table 2.2, has risen steadily from 1987 to 2012.

Table 2.2: Economy-wide Concentration, Sales Share of Top-50 Firms

Year	1987	1992	1997	2002	2007	2012
Share50	26.00%	26.57	26.69	29.52	30.51	31.31

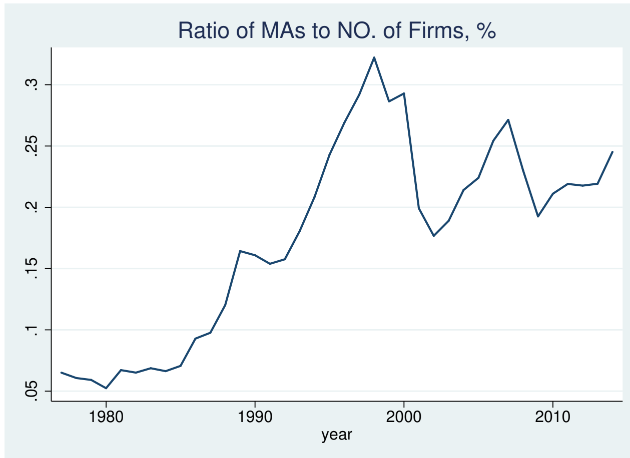
Data source: Boldrin and Zhu (2018)

2.2.2 The Rising M&A Activities

On the other hand, since the Reagan administration took office in the early 1980s, antitrust policy implementation has been relaxed drastically (Mueller, 1984). A more lenient antitrust policy is reflected in a rising Mergers and acquisitions cases after the 1980s. Figure 2.3 presents the ratio of total M&As to total NO of firms in the U.S. from 1977 to 2015[10]. This ratio was 0.065% in 1977, relatively stable in the late 1970s and early 1980s, and has risen to 0.245% in 2014. M&A activity in Figure 2.3 demonstrates a clear up-and-down wave patterns. In current paper, we focus on the long run trend from the 1980s on wards, and do not address this wave pattern. Figure 2.7 in appendix shows total number of M&A from 1850 to present. The post-1980s period stands out in even longer historical horizons.

[10]Total value of M&As increases from 306 trillion U.S. dollars in 1985 to 1741 in 2017, an increase of 469%.

Figure 2.3: Ratio of M&As to Number of Firms, 1977-2015



Data source: https://imaa-institute.org.

Market power is an important consideration in antitrust policy implementation. Empirically, firm's M&A activity is expected to be closely related to its markup. To test the effect of M&A on markup, we first merge Compustat firm data with SDC-M&A data, which starts in 1980. From 1980 to 2015, there are $246,452$ *firm* $\times$ *year* observations in Compustat[11]. Among which $40,291$, or 16%, have at least one M&A.

Using the combined data, we run the following regression

$$\ln(\frac{markup_{i,t+1}}{markup_{i,t}}) = \beta_0 + \beta_1 MA_{i,t} + \beta_2 \ln(markup_{i,t}) + controls + \epsilon,$$

where $MA_{i,t}$ is a dummy variable and equals to 1 if firm i has at least one mergers & acquisitions in year t. The dependent variable is the relative change in markup for firm i from year t to year $t+1$. Firm's current markup is added as a control as relative change is expected to be small given a firm already have a large value of markup today. Year and sector dummies are also added as controls in the regression. Table 2.3 presents the regression

[11] This number is counted after dropping observations with missing values.

results

Table 2.3: Effect of M&A on Markup

Dep. var.: $\Delta \ln(markup)$	(1)	(2)
$\ln(markup)$	-0.240***	-0.246***
	(0.001)	(0.001)
I^{MA}	0.019***	0.001
	(0.003)	(0.003)
$\ln(markup) \times I^{MA}$		0.06***
		(0.005)
Year D	Yes	Yes
Sector D	Yes	Yes
R^2	0.12	0.12
NO	221,213	221,213

Date source: Compustat and SDC-MA

Column 1 shows the baseline results. The coefficient in front of I_{MA} is positive and significant. The effect is economically sound, firm's markup increases 1.9% in average after an M&A activity. Column 2 adds the interaction term between markup and the M&A dummy, to capture the conjecture that the amplification effect of M&A on markup might be larger for firms which have already had a larger markup. Year and sector dummies are also added. As in column 1, the coefficient in front of the interaction term is positive and significant.

2.3 Model

We are now ready to provide a modeling framework to evaluate costs and benefits of antitrust policy. The model here builds on David (2017), which is a firm dynamics model with Mergers and acquisitions. Firms are heterogeneously endowed with a productivity, which can be changed through M&A and is otherwise constant. Upon paying a cost, firms search for potential targets/acquirers and successfully matched pairs decide whether to consummate a M&A deal. Due to technical complementarity between acquirers and targets, acquirers increase their productivity by merging other firms, with the magnitude of increase depends on the pre-merger size of targets. We introduce heterogeneous markup and international trade into that framework. In particular, heterogeneous markup is modeled as a result of different elasticity of demand faced by different firms. We start with a closed economy version of the model and then extend it to an open economy.

2.3.1 Benchmark Economy

The economy consists of two sectors, an intermediate goods sector where each single firm monopolies in producing an intermediate variety, and a final good sector where firms produce competitively by aggregating over intermediate goods. We start with the competitive final good sector.

Final Good The final good is produced competitively by aggregating over intermediate goods according to the following aggregator

$$\int \sigma\left(\frac{y_t(\omega)}{Y_t}\right) d\omega = 1,$$

where Y_t denotes quantity of the final good in period t and $y_t(\omega)$ intermediate variety ω. Following Kimball (1995), $\sigma(x)$ is a strictly increasing and strictly concave function, satisfying $\sigma(1) = 1$. The case where $\sigma(x) = x^{\frac{\epsilon-1}{\epsilon}}$ gives the typical CES aggregator.

Normalize the price of final good as 1 and denote $p_t(\omega)$ the price of intermediate goods ω. Final good producers optimally choose intermediate varieties to maximize profit, i.e.

$$\max_{y_t(\omega)} \quad Y_t - \int p_t(\omega) y_t(\omega) d\omega,$$

subject to the production function above. The maximization problem gives the demand function facing the producer of intermediate variety ω as

$$p_t(\omega) = \sigma' \left(\frac{y_t(\omega)}{Y_t} \right) D_t,$$

where D_t is the common term facing all intermediate goods producers and defined as

$$D_t \equiv \left(\int \sigma' \left(\frac{y_t(\omega)}{Y_t} \right) \frac{y_t(\omega)}{Y_t} d\omega \right)^{-1}.$$

One property of the derived demand function facing intermediate varieties is that, depending on the specific functional form of $\sigma(x)$, the elasticity of demand might vary across intermediate goods producing firms. This generates heterogeneous markups among intermediate firms.

Intermediate Goods Each intermediate good is produced by a single firm. An intermediate firm with productivity z uses labor as the only input and accesses to the following linear production technology

$$y_t = z\ell_t^{\delta}, \quad \delta \in (0, 1].$$

Intermediate firms optimally choose quantity and price of its products and the amount of labor to employ, ℓ, to maximize current period profit

$$\pi_t(z) \equiv \max_{p_t(\omega), y_t(\omega), \ell_t(\omega)} p_t(\omega) y_t(\omega) - W_t \ell_t(\omega),$$

where W_t is common wage rate in period t. Intermediate firms are subject to the demand scheme from the final goods sector.

As in Klenow and Willis (2016) and Edmond, Midrigan and Xu (2018), we choose the following specification of the kimball aggregator

$$\sigma(q) = 1 + (\beta - 1)\exp\left(\frac{1}{\alpha}\right)\alpha^{\frac{\beta}{\alpha}-1}\left[\Gamma\left(\frac{\beta}{\alpha}, \frac{1}{\alpha}\right) - \Gamma\left(\frac{\beta}{\alpha}, \frac{q^{\alpha/\beta}}{\alpha}\right)\right],$$

where $\alpha \geq 0$ and $\beta > 1$, and $\Gamma(a, b) \equiv \int_b^\infty x^{\alpha-1}e^{-x}dx$ is the upper incomplete Gamma function. Under this aggregation, the term $\sigma'(q)$, with $q \equiv \frac{y}{Y}$ representing the relative size of the intermediate firm, in the intermediate goods demand function, is given by

$$\sigma'(q) = \frac{\beta - 1}{\beta}\exp\left(\frac{1 - q^{\alpha/\beta}}{\alpha}\right),$$

which is a decreasing function of relative size. That is, large firm (i.e. with larger q) faces a smaller demand elasticity and optimally charge a higher markup. To see this, use this demand function to substitute away price in intermediate firms' profit maximization problem and rewrite it as

$$\begin{aligned}\pi(z) &\equiv py - W\ell \\ &= \max_q \left[\sigma'(q)q - \frac{WY^{\frac{1}{\delta}-1}}{Dz^{\frac{1}{\delta}}}q^{\frac{1}{\delta}}\right]DY,\end{aligned}$$

where we use the fact that in equilibrium, $\ell = \left(\frac{y}{z}\right)^{\frac{1}{\delta}}$. The optimal condition with respect to q yields

$$\sigma'(q)\left(\frac{\beta - q^{\alpha/\beta}}{\beta}\right) = \frac{WY^{\frac{1}{\delta}-1}}{Dz^{\frac{1}{\delta}}}\frac{q^{\frac{1}{\delta}-1}}{\delta}.$$

From this first order condition, we can easily verify that $q^*(z)$ is strictly increasing in z. Firms with a higher productivity have a larger market shares. The markup, i.e. the ratio of price of marginal cost, satisfies

$$\mathcal{M} \equiv \frac{p}{MC} = \frac{\beta}{\beta - (q^*)^{\alpha/\beta}}.$$

That is, markup is an increasing function of q. larger firms charge a higher markup. The profit of an intermediate firm is

$$\pi(z) = \sigma'(q^*)\frac{(q^*)^{(\alpha+\beta)/\beta}}{\beta}DY,$$

which is strictly increasing in q^* in the interval of $[0, (\alpha+\beta)^{\beta/\alpha}]$. Since $q^*(z) \leq \beta^{\beta/\alpha}, \forall z$, the profit $\pi(z)$ is also strictly increasing in z, although we no longer have a simple closed form of it as in the CES case ($\alpha = 0$).

Mergers and Acquisitions To single out the effect of M&A, we assume that intermediate firms' productivity and size can change after M&A, and stay constant otherwise. Following David (2017), M&A is modeled under a search and matching framework. Each firm simultaneously decide to search to be an acquirer or/and a target in the M&A market. Upon successfully match with a target which has a productivity z_t, the post-merger productivity of an acquirer firm with pre-merger productivity z_a is

$$z_m = m(z_a, z_t).$$

It is assumed that m is an increasing function in both of its arguments. The function captures the productivity increase due to technological complementarity between acquirers and targets. Denote $V(z)$ the value of firm with productivity z, which will be specified below. M&A possibly increases acquirer's size and value, and the potential gain from M&A is

$$G(z_a, z_t) = V(z_m) - V(z_a) - V(z_t),$$

and a M&A will be proposed if and only if $G(z_a, z_t) > 0$.

A M&A proposal has a probability $\tau(z_a, z_t) \in [0, 1]$ of being passed by the policy maker. τ is allowed to be dependent on (z_a, z_t). If a M&A is proposed and passed, the surplus is split between the acquirer and the target through Nash Bargaining, denote γ as the bargaining power of the acquirer. Define G_a and G_t as the expected net gains for the acquirer and target, from a successful matching, respectively. It follows that

$$G_a(z_a, z_t) = \max\Big\{\tau(z_a, z_t)\gamma G(z_a, z_t), 0\Big\};$$
$$G_t(z_a, z_t) = \max\Big\{\tau(z_a, z_t)(1-\gamma)G(z_a, z_t), 0\Big\}.$$

In a consummated M&A, the acquirer pays the target

$$P(z_a, z_t) = V(z_t) + (1-\gamma)G(z_a, z_t),$$

and continues with the post-merger productivity z_m, while the target exits the economy. If the M&A does not happen, both firms continues with their original productivity. We omit the possible financial constraints, and assume that an acquirer can always afford to pay $P(z_a, z_t)$, either by a lump-sum transfer or a long-term contract.

As in David (2017), firms simultaneously choose search intensities $\lambda(z)$ of meeting a potential target and $\mu(z)$ of meeting a potential acquirer, with the associated costs given by

$$c(x;\phi), \quad \textit{for} \quad x = \lambda, \mu.$$

$c(x;\phi)$ is increasing in x, implying a higher cost for a higher search intensity, and also increases with ϕ, which represents the strength of antitrust policy. A higher value of ϕ implies a larger M&A cost, and therefore represents a stricter antitrust policy. A relaxation of antitrust policy is modeled as a decrease in the value of ϕ.

Market tightness on the acquirer (target) side is defined as the total search intensity from potential targets (acquirers) to that from potential acquirers (targets). A higher tightness implies a higher matching rate. Denote $F(z)$ the distribution of active firms, which will be specified below. The market tightness on the acquirer and target side is[12]

$$\theta_a = \frac{\int \mu(z)dF(z)}{\int \lambda(z)dF(z)}, \quad \theta_t = \frac{\int \lambda(z)dF(z)}{\int \mu(z)dF(z)}.$$

The rate a type z_a acquirer meets a type z_t target, and vice versa, is given by

$$\lambda(z_a)\theta_a \underbrace{\frac{\mu(z_t)dF(z_t)}{\int \mu(z)dF(z)}}_{\Omega(z_t)}; \quad \mu(z_a)\theta_t \underbrace{\frac{\lambda(z_a)dF(z_a)}{\int \lambda(z)dF(z)}}_{\Phi(z_a)}.$$

where $\lambda(z_a)\theta_a$ denotes the rate the acquirer z_a meets a target, and $\Omega(z_t)$ denotes the conditional probability that this target has productivity z_t. A symmetric interpretation applies on the target side.

Value Functions Assume that firms exit at an exogenous rate η. The value of a firm as a function of the state variable, productivity z, is

$$(\rho + \eta)V(z) = \max_{\lambda,\mu} \; \pi(z) - c(\lambda;\phi) - c(\mu;\phi) + \lambda\theta_a \mathbb{E}_{z_t}[G_a(z, z_t)] + \mu\theta_t \mathbb{E}_{z_a}[G_t(z_a, z)].$$

Firms optimally choose search intensities, both as an acquirer and a target[13]. The flow value of a firm with productivity z is equal to, the instantaneous profit π, plus the expected net benefit from searching in the M&A market.

Entry, Exit and Equilibrium As mentioned earlier, firms exit with an exogenous rate η. A firm enters the economy by paying a fixed cost c_e, then draws its initial productivity z from an exogenous distribution $H(z)$. Free entry implies the expected value of entry equals this fixed entry cost, i.e.

$$\int V(z)dH(z) = c_e.$$

[12]Note that as the model is in continuous time, these two tightness index do not have to be smaller than 1.

[13]In an infinitely small interval of time, the probability of simultaneously meeting a potential target and a potential acquirer is zero.

Denote M_e total mass of entrants, and M mass of all active firms. The distribution of firm productivity, $F(z)$, evolves according to the following Kolmogorov forward equation.

$$
\begin{aligned}
d\dot{F}(z) = & \underbrace{\int \lambda(z_a)\theta_a \Big[I\Big(G\big(z_a, m^{-1}(z, z_a)\big) > 0\Big)\Omega\Big(m^{-1}(z, z_a)\Big)\Big] dF(z_a)}_{\textit{inflow through M\&A}} + \underbrace{\frac{M_e}{M} dH(z)}_{\textit{inflow through entry}} \\
& - \underbrace{\lambda(z)\theta_a dF(z) \int I(G(z, z_t) > 0)\Omega(z_t)}_{\textit{outflow through merging}} - \underbrace{\mu(z)\theta_t dF(z) \int I(G(z_a, z) > 0)\Phi(z_a)}_{\textit{outflow through being merged}} \\
& - \underbrace{\eta dF(z)}_{\textit{outflow through exog. exit}} .
\end{aligned}
$$

This law of motion decompose changes in probability over any state z into 5 possible channels: Firms with a different productivity might merger other firms and obtain a post-merger productivity z; Exogenous entrants might draw a productivity z; Firms originally with productivity z might merger other firms and arrive at a new (higher) productivity; Firms originally with productivity z might be merged by other firms and exit the economy; Firms originally with productivity z might exit upon being hit by the exogenous exit shock. In a stationary equilibrium, the mass at each state is constant. That is, change is equal to zero,

$$d\dot{F}(z) = 0, \quad \forall\, z.$$

The economy admits a representative household, who owns all firms, and wage and firm profits. It also pays the entry cost c_e. The household consumes all it income every period as saving is not allowed. It is endowed with 1 unit of labor, and supplies it inelastically.

We close the model with market clearing conditions. Labor is used in producing intermediate goods, all employed workers in the intermediate goods sector should equal to total supply, i.e.

$$M \int \left(\frac{y(z)}{z}\right)^{\frac{1}{\delta}} dF(z) = 1.$$

The final good can be used in three ways: consumption, payment of search cost in the M&A market, and payment of entry cost. Market clearing implies

$$Y = C + Y_s + M_e C_e,$$

where total search cost equals that for acquirers and targets. That is

$$Y_s = M \int c\left(\lambda(z); \phi\right) dF(z) + M \int c\left(\mu(z); \phi\right) dF(z).$$

Now we can formally define the equilibrium in our model.

Definition 2. *We focus on a **S**tationary **E**quilibrium of the economy, which in the closed economy version of model consists of: aggregate variables,* $\Big\{Y, W, C, M, M_e, F(z)\Big\}$*; Firm's policy function and profit function from the static profit maximization problem,* $\Big\{q(z), \pi(z)\Big\}$*; Firm's search intensity policy functions and value function from the dynamic search and matching problem,* $\Big\{\lambda(z), \mu(z), V(z)\Big\}$*, such that*

1. *Policy functions solve their corresponding maximization problems;*

2. *The free entry condition is met;*

3. *The goods and labor markets clear;*

4. *The evolution of firm characteristic distribution is consistent with the stationary conditions.*

2.3.2 Trade and New Entrants' Quality

We extend the model into an open economy framework in the following way. Now with trade, assume the production function of the final good becomes

$$\int \sigma \left(\frac{y_t(\omega) + y_t^F(\omega)}{Y_t} \right) d\omega = 1,$$

where $y_t^F(\omega) > 0$ implies net import, and $y_t^F < 0$ implies net export. $y_t(\omega)$ is total amount of intermediate variety ω. In equilibrium, it always holds that $y_t^*(\omega) + y_t^F(\omega) > 0$. To simplify analysis, $p_t^F(\omega)$ are set to the same as $p_t(\omega)$.

Now the intermediate good demand function is

$$p_t(\omega) = \sigma' \left(\frac{y_t(\omega) + y_t^F(\omega)}{Y_t} \right) D_t.$$

Intermediate firms' profit maximization problem becomes

$$\pi(z) = \max_q \left[\sigma' \left(\max \left(q + q^F(z), 0 \right) \right) q - \frac{WY^{\frac{1}{\delta}-1}}{Dz^{\frac{1}{\delta}}} q^{\frac{1}{\delta}} \right] DY,$$

where $q^F(z)$ is exogenously given.[14]

With a correctly specified marginal cost $W/D < \sigma' \left(max \left(q^F(z), 0 \right) \right)$, we could reach interior solutions for all z. The implied markup function is

$$\mathcal{M} = \frac{p}{MC} = \frac{\sigma'(q^* + q^F)}{\sigma'(q^* + q^F) + \sigma''(q^* + q^F) q^*}.$$

The M&A, entry, exit, stationary distribution, and the labor market clearing condition do not change. The only change comes from the final good market clearing condition, which

[14]Note that when $q^F < 0$, the concavity of the maximization problem above could be violated locally

now is

$$Y = C + Y_s + M_e C_e + Y^F,$$

where C, Y_s and M_e are defined as in the closed economy, and

$$Y^F = M \int p^F(z) y^F(z) dF(z)$$

is payment to (from) foreign market, with

$$p^F(z) = p(z) = \sigma' \left(q^*(z) + q^F(z)\right) D.$$

2.4 Quantitative Analysis

This section provides the quantitative results. We first specify a few function forms. The merger technology is set as

$$z_m = A z_a^\kappa z_t^\epsilon$$

where $A > 0$, $\kappa > \epsilon > 0$. The latter implies that post-merger productivity relies more heavily on the productivity of acquirers, and it is therefore more likely that large firms acquire small ones. The search cost function is chosen to be

$$c(x; \phi) = \frac{\phi^{\nu-1}}{\nu} x^\nu$$

where $\phi > 0$, $\nu > 1$. As detailed in the model part, the Kimball aggregator is set to be the Klenow-Willis form.

Parameters There are 12 parameters in the model. For M&A technology related parameters, A, κ and ϵ, we borrow the estimated values from David (2017), and for parameters in the kimball aggregator from Edmond, Midrigan, and Xu (2018). We pick the discount rate to be 0.05 to match an annual interest rate of 5%, and η to be 0.04 to target an annual exit rate of 4%. δ, which governs the decreasing return to scale in intermediate goods production, is chosen to be 0.8 to match a 20% of profit rate. The entry cost c_e is normalized to 1.[15] Table 2.4 summarizes the current parameterization

Table 2.4: Parameterization

Para.	meaning	values	source
ρ	discount rate	0.05	interest rate
δ	para. in intermediate prod. func.	0.8	20% profit rate
η	exit rate	0.04	
γ	bargaining power	0.49	
ξ	para. in new entrant's prod. distr.	10	
A	scale para. in M&A tech.	1.05	David'17
κ	acquirer productivity elasticity	0.91	David'17
ϵ	target productivity elasticity	0.53	David'17
ν	para. in search cost func.	8	
c_e	entry cost	1	normalization
α	para. in kimball aggr.	2.18	EMX'18
β	para. in kimball aggr.	11.55	EMX'18

We pick ϕ and τ as "technology parameter" and "policy parameter", and try to pin down their optimal values. We then experiment on different c_e and q^F, try to figure out their effects on the optimal parameter values. We hope this exercise would shed light on why the antitrust policy has been weakened since the 1980s.

Under the benchmark parameterization and set $\tau = 1$, household welfare shows a clear hump-shaped curve as we vary ϕ.

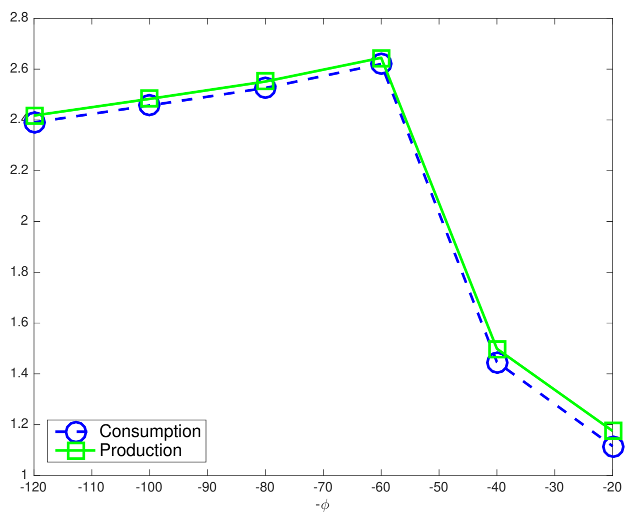
[15] We are currently working on a more deliberate calibration.

Table 2.5: Welfare under Different ϕ

ϕ	10	20	40	60	80	100
C	1.7334	1.7777	2.1741	2.2072	2.1202	2.0849
Y	1.9171	1.9478	2.2603	2.2811	2.2119	2.1830

This implies an optimal technology parameter value at $\phi^* = 60$.[16] Figure 2.4 plots consumption and output under different values of ϕ.

Figure 2.4: Consumption and Output under Different ϕ



Open Economy The next part presents results we have experimented with several versions of the $q^F(z)$ function. If $q^F(z) > 0$ (< 0), the economy is net importing (exporting) in all varieties with characteristic z. We keep $c_e = 1$ in those exercises.

[16]Unfortunately, we don't observe such pattern when we keep ϕ constant and vary τ. So we pick ϕ as our 'parameter of interest' and do comparative statics.

Case 1: $q^F(z) = x^{index} - 1$

x	1.00	1.02	1.05	1.08	1.10	1.15	1.20
ϕ^*	60	60	60	60	60	60	40
Y^F/Y	0%	7.34	17.70	26.76	32.63	48.42	69.43

Case 2: $q^F(z) = x^z - 1$

x	1.00	1.10	1.15	1.20
ϕ^*	60	40	40	40
Y^F/Y	0%	19.05	26.35	31.47

Case 3: $q^F(z) = z^x - yz$

x, y	1.00 , 1.00	1.30 , 0.70	1.50 , 0.90	2.00 , 0.95
ϕ^*	60	40	60	40
Y^F/Y	0%	42.60	32.80	52.87

We have then done an exercise where international trade has been shut down, but the quality of entrants is allowed to vary. We change z_{peak} of $H(z)$, while keep it as a Pareto distribution on $[z_{peak}, \infty]$. The following table presents the results.

Table 2.6: Entry Quality and Optimal ϕ

z_{peak}	1.00	1.10	1.21	1.32	1.45
ϕ^*	60	60	80	120	140

As the new entrant firm's productivity distribution shifts to the right, the optimal ϕ^* increases. If the productivity of young firms relative to old incumbent firms decline over time, the implied optimal antitrust policy should be more lenient.

2.5 Conclusion

This paper provides a framework to evaluate the costs and benefits of antitrust policy in a dynamic general equilibrium framework. We first provide empirical evidence for evolving distribution of markup across public firms and the effect of mergers and acquisitions on acquiring firms' markup. We then present a dynamic general equilibrium model that incorporates both the productivity enhancing and markup increasing effect of mergers and acquisitions. Optimal antitrust policy seeks a balance between these two forces. We showed that in a closed economy version of our model, Welfare, measured as aggregate consumption in a stationary equilibrium, is a hump-shaped function of antitrust policy.

We extend the benchmark model to an open economy and aim to formalize the following intuition: in a globalized economy, increasing productivity fully accrues to domestic firms while a higher markup only partially hurts domestic consumers. A weakening antitrust policy since the 1980s is thus an optimal response to the increasing globalization in the same period.

2.6 Appendix

2.6.1 Measurement of Markups

We use the price-marginal cost markup, $\frac{P_i}{MC_i}$, to indicate the monopoly power of firm i, and measure it by using methods proposed in De Loecker and Warzynski (2012) and De Loecker and Eeckhout (2017). The method starts with the typical cost minimization problem of firm i is[17]

$$\mathcal{L}(V_i, K_i, \lambda_i) = P_i^V V_i + r_i K_i - \lambda_i \left(F(V_i, K_i) - \bar{Y} \right),$$

[17] Note the implicit assumption here is that

where P_i^V and V_i are the price and quantity of the variable input, r_i and K_i are the user cost and quantity of capital. $F(.)$ is the production function and $\bar{Y}$ is the targeted output level, and λ_i is the Lagrangian multiplier. The advantage of writing in this way is that $\lambda = \frac{\partial \mathcal{L}}{\partial \bar{Y}}$, i.e. λ_i gives the marginal cost of firm i. The first order condition *w.r.t.* V_i reads

$$P_i^V - \lambda_i \frac{\partial Y_i}{\partial V_i} = 0.$$

$Y_i \equiv F(V_i, K_i)$ denotes total output. Equivalently

$$\epsilon_i^V \equiv \frac{\partial Y_i}{\partial V_i} \frac{V_i}{Y_i} = \frac{1}{\lambda_i} \frac{P_i^V V_i}{Y_i}.$$

It follows that markup, $\mu_i = \frac{P_i}{\lambda_i}$, equals to

$$\mu_i = \epsilon_i^V \frac{P_i Y_i}{P_i^V V_i},$$

where $P_i Y_i$ and $P_i^V V_i$ are observed in data, and $\epsilon_i^V \equiv \frac{\partial Y_i}{\partial V_i} \frac{V_i}{Y_i}$ is the elasticity of output *w.r.t.* the variable input and can be estimated from data.

Assume the production function is, $Q = F(V, K)\exp(z)$[18]. Take log on both sides and use lower case letters to denote natural logarithmic of variables,

$$q = \xi_v v + \xi_k k + z + \epsilon.$$

Demand of the variable input, V, is a function of capital stock, K, and unobserved productivity, Z, V=g(K, Z). We can then represent productivity, $Z(z)$, as a function of $K(k)$ and $V(v)$, i.e., $z = h(k, v)$. In the first stage, run the following regression non-parametrically (or

[18]We assume the same production function at 2-digit NAICS sector level among Compustat firms.

approximate ϕ by a polynomial)

$$q = \phi(v, k) + \epsilon.$$

Further assume the exogenous productivity follows a $AR(1)$ process

$$z_{t+1} = \rho z_t + \epsilon_z.$$

Obtain $\hat{z} = \hat{\phi}(v, k) - \xi_v v - \xi_k k$ from the first stage, and we then apply general method of moments to estimate ξ_v, by using the following moment conditions:

$$E\left[(\hat{z}_t - \rho \hat{z}_{t-1}) X_{t-1}\right] = 0.$$

In the baseline case, we include capital stock in period t, k_t, and variable input in period $t-1$, v_{t-1}, in X_{t-1}.

2.6.2 Solution Algorithm

- Guess
  - Loop1: Guess $w \equiv \frac{WY^{\frac{1}{\delta}-1}}{D}$, solve for $q^*(z)$ [19] from

    $$\max_{q \geq max\{0, -q^F\}} \quad \sigma'\left(q + q^F\right) q - \frac{w}{z^{\frac{1}{\delta}}} q^{\frac{1}{\delta}}$$

    whose f.o.c. is:

    $$\sigma'\left(q^* + q^F\right) + \sigma''\left(q^* + q^F\right) q^* = \frac{w}{\delta z^{\frac{1}{\delta}}} \left(q^*\right)^{\frac{1}{\delta}-1}$$

[19] As we see in the model part, equilibrium q^* will surely such that $q^* + q^F \geq 0$.

– Loop2: Guess $dF(z)$, solve for $M, D, Y, W, \pi(z)$ from

$$M = \left(\int \sigma \left(q^* + q^F \right) dF(z) \right)^{-1}$$

$$D = \left(M \int \sigma' \left(q^* + q^F \right) \left(q^* + q^F \right) dF(z) \right)^{-1}$$

$$Y = \left(M \int \left(\frac{q^*}{z} \right)^{\frac{1}{\delta}} dF(z) \right)^{-\delta}$$

$$W = \frac{wD}{Y^{\frac{1}{\delta}-1}}$$

$$\pi(z) = \left[\sigma' \left(q^* + q^F \right) q^* - \frac{w}{z^{\frac{1}{\delta}}} \left(q^* \right)^{\frac{1}{\delta}} \right] DY$$

– Loop3: Guess $V(z)$, construct $G(z_a, z_t)$ matrix based on $V(z)$.

– Loop4: Guess $\mu(z)$ and θ_a, solve for $\lambda(z)$ and θ_t from (2).

- Update

  – Loop4: Use $\lambda(z)$ and θ_t to solve for new $\mu(z)$ and θ_a from (2), update till converge.

  – Loop3: Use $\mu(z)$, $\lambda(z)$, θ_a, θ_t to get a new $V(z)$ from (1), update till converge and choose step small.

  – Loop2: Use $G(z_a, z_t)$, $\mu(z)$, $\lambda(z)$, θ_a, θ_t and M to solve for M_e from:

$$M_e = M\eta + M \int \mu(z_t) \theta_t \left[\int \tau(z_a, z_t) I(G(z_a, z_t)) \Phi(z_a) \right] dF(z_t)$$

With the exogenous $dH(z)$, we can use the KFE to get a new

$$dF_{+1}(z) = dF(z) + step \cdot (RHS\ of\ KFE)$$

update till converge.

– Loop1: Check the free entry condition, if value>cost, increase w, otherwise reduce it, till free entry condition is met or the change is too small.

2.6.3 Tables and Figures

Table 2.7: Number of Firms in Compustat Sample

Year	NO	Year	NO	Year	NO	Year	NO
1951	441	1968	3015	1985	5929	2002	7299
1952	456	1969	3623	1986	6242	2003	7048
1953	474	1970	3717	1987	6430	2004	6873
1954	486	1971	3768	1988	6369	2005	6680
1955	516	1972	3885	1989	6231	2006	6553
1956	553	1973	3956	1990	6203	2007	6309
1957	573	1974	4354	1991	6293	2008	6027
1958	608	1975	5785	1992	6502	2009	5972
1959	637	1976	5804	1993	6867	2010	5757
1960	663	1977	5776	1994	7269	2011	5607
1961	1281	1978	5672	1995	7632	2012	5543
1962	1667	1979	5495	1996	8487	2013	5758
1963	1934	1980	5381	1997	8534	2014	5712
1964	2183	1981	5432	1998	8207	2015	5458
1965	2435	1982	5421	1999	8453	2016	5217
1966	2620	1983	5810	2000	8234	2017	4915
1967	2819	1984	5950	2001	7797		

Data source: Compustat

Table 2.8: Concentration in 2-digit Sectors

	Share50, 2-digit				Share04, 6-digit average			
	1997	2002	2007	2012	1997	2002	2007	2012
Wholesale Trade	20.3%	27.2	24.9	27.6	24.3%	31.4	29.1	30.8
Retail Trade	25.7	31.7	33.3	36.9	18.5	26.8	31.0	34.6
Transportation	30.7	33.0	42.7	42.1	23.6	24.5	30.8	35.2
Utilities	64.5	69	70.1	69.1	25.8	23.2	23.0	24.2
Information		62	62	62.3		49.8	52.4	52.0
Finance	38.6	44.9	46	48.5	26.0	32.0	36.1	35.4
Real Estate	19.5	24.4	26.1	24.9	18.8	24.0	25.1	23.3
Prof. Sci. Tech.	16.2	16.5	18.6	19.0	15.0	15.1	17.9	18.3
Administrative	22.1	21.9	23.0	23.7	21.8	23.1	24.4	24.4
Education	19.6	23.2	23.5	23.8	16.6	19.4	19.4	19.4
Health Care	18.8	17.2	17.4	19.6	16.2	15.1	15.1	17.0
Entertainment	21.8	23.5	24.1	24.3	19.2	20.5	21.5	21.6
Accommodation	21.1	23.1	23.7	21.2	13.8	17.4	18.5	16.0
Other Services	12.8	14	13.8	12.6	13.8	14.7	15.0	14.1

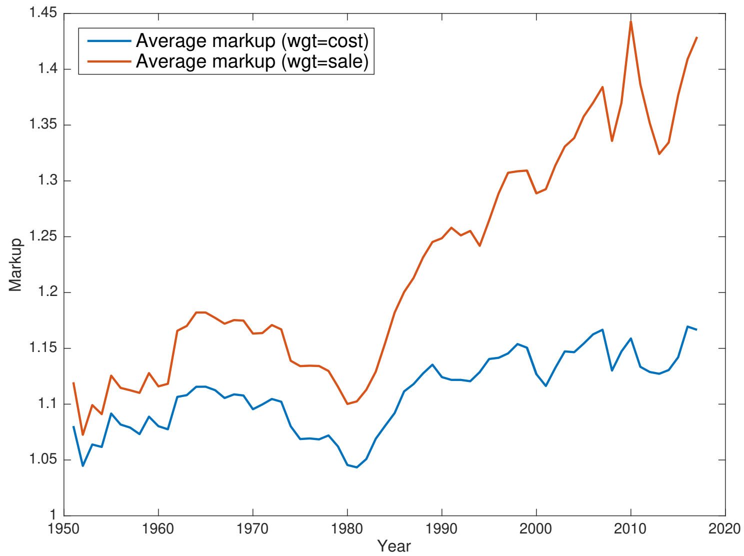
Note: For most services sectors in 1997, statistics are only available for establishments subject to federal income taxes (instead of all establishments) in Service sectors. To be consistent, the same criteria is applied to 2002, 2007, and 2012.

Table 2.9: Total Number of Firms in the Economy (Unit: thousand)

Year	1977	1982	1987	1992	1997	2002	2007	2012
Economy	3147.9	3604.0	4179.8	4377.1	4752.3	4908.7	5240.0	4979.5
MFG	261.2	272.9	290.8	296.0	303.2	283.4	267.8	234.4
WHO	277.9	307.6	337.0	354.8	372.9	349.6	341.4	310.8
RET	942.8	912.7	953.0	939.8	955.6	949.5	980.0	953.0
TCP	121.3	130.5	153.9	162.2	187.9	190.1	195.6	185.4
FIRE	298.0	299.5	347.2	358.1	393.7	429.8	489.7	435.9
SRV	1122.5	1288.5	1600.8	1741.6	1924.9	2055.0	2344.1	2355.5

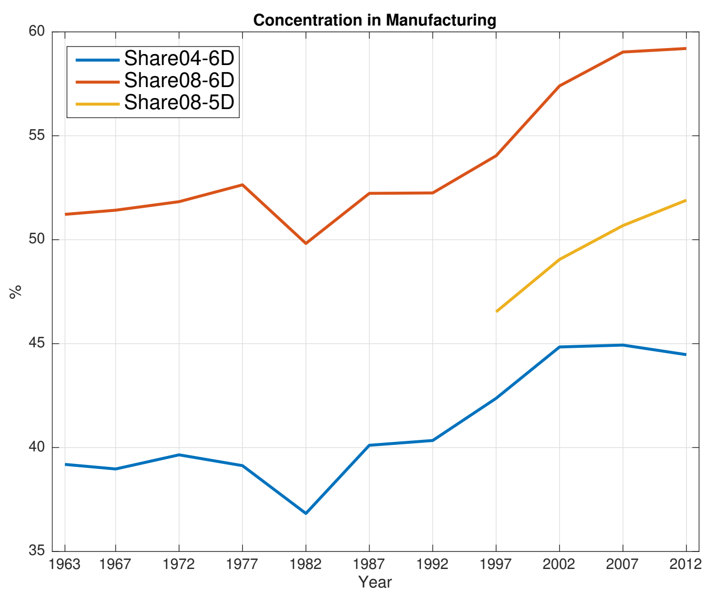
Note: MFG-Manufacturing; WHO-Wholesale trade; RET-Retail trade;
TCP-Transportation, communication and public utilities; SRV-Services.
Data source: Business Dynamics Statistics

Figure 2.5: Average Markups, 1950-2017



Note: Average markups among public firms, with weights equal to cost of goods sold for the blue line and sales for the red line. *Data source:* Compustat.

Figure 2.6: Concentration in Manufacturing, 1963-2012



Note: The blue (resp. red) line plots the average revenue share of the 4 (resp. 8) largest firms across 6-digit manufacturing sectors; the yellow line is the average revenue share of the 8 largest firms across 5-digit manufacturing sectors. All values are weighted by revenue.

Figure 2.7: Number of Mergers & Acquisitions in the U.S., 1850-2010

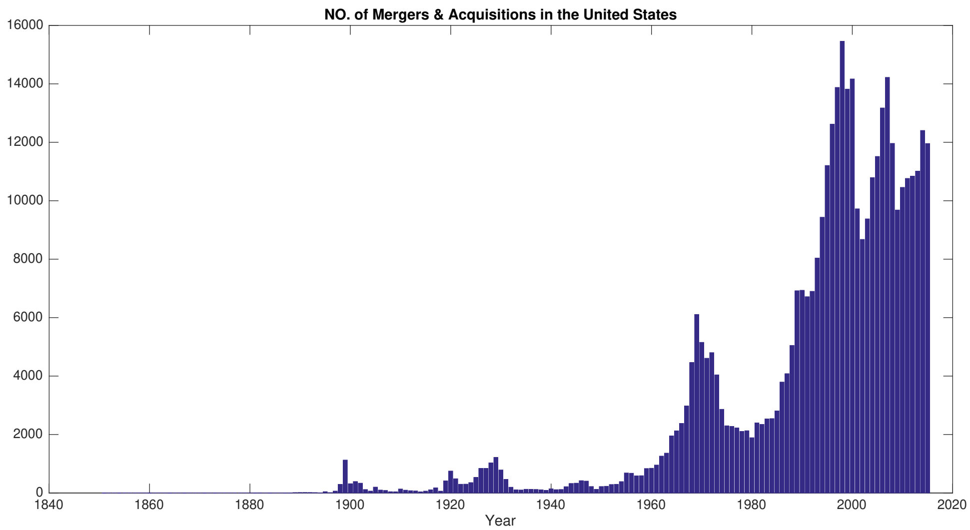
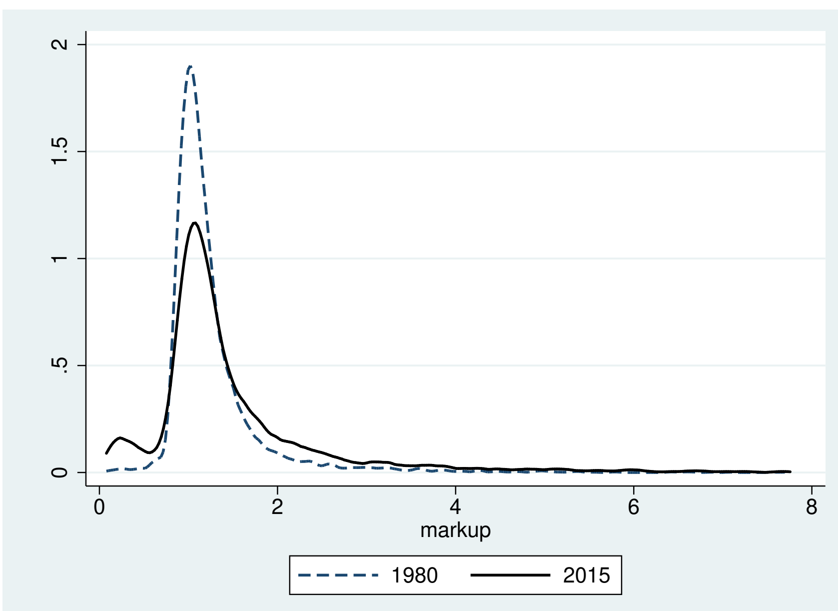


Figure 2.8: Distribution of Markups among Public Firms in 1980 and 2015, 1-99 Percentiles



Note: Distribution of markups conditional on markup values between 1 and 99 percentiles of the whole distribution.*Data source:* Compustat.

Anecdotal Evidence from Court Reports Reports from Mergers & Acquisitions law cases suggest that international competition is an important consideration in antitrust policy implementation. One such case is the Boeing/McDonnell Douglas merger in the 1990s[20]. At that time, Boeing was the largest producer and accounts for about 64% of sales in the world commercial jet aircraft market. The other two producers were Airbus Industries, a European manufacturer which accounts for about 30%, and MaDonnell Douglas from the U.S., with about 5%. Boeing submitted petition to both U.S. and the European Commission. It was approved by the Americans and nearly prohibited by the European Union, which eventually clear the merger after imposing strong conditions.

Both the U.S. Federal Trade Commission (FTC) and the European Commission investigated the case, but they reached very different conclusion. The FTC has concluded that the acquisition of McDonnell Douglass by Boeing would not create a monopoly or substantially lessen competition in the commercial aircraft market. However, the European Commission concluded that the merger would increase Boeing's dominance. In its study, it was found that the McDonnell Douglass' presence led to a reduction of over 7% in the realized price. After the FTC investigation, the Clinton Administration argued to key European official that the merger is not anti-competitive and important to the employment in the United States, and even threatened to retaliate if Europe undermines the merger. The European Commission eventually backed away from a prohibition of the merger, but imposed significant conditions, such as Boeing not to enforce its exclusivity rights under the agreements with big American airline companies.

[20] This case is documented in Eleanor M. Fox (2012).

Chapter 3

A Model of Technology Diffusion

Linyi Cao[1] Cheng Wang[2]

3.1 Model

Let time be discrete and denoted $t = 0, 1, ...$ and there is one good in the model economy. An industry consists of a continuum of producers, normalized to be of measure one. These producers have been in the industry for a long time, and up to time $t = 0$ all of them have used an incumbent technology to produce θ_0 (> 0) units of the good each period, with certainty. But at $t = 0$ a new technology has come to existence, ready to be adopted by all producers in the industry. This new technology is supposed to be more efficient relative to the old but, being new, also brings with it uncertainty about how reliable it is.

The new technology may be reliable as desired, in which case it produces a high output of θ_H ($> \theta_0$) units of the good with certainty. It could also be unreliable, in which case it produces either the output θ_H or an output θ_L (< 0).

[1] Department of Economics, Washington University in St. Louis, One Brookings Drive, St. Louis, MO 63130 (email: *l.cao@wustl.edu*).

[2] School of Economics, Fudan University (email: *wangcheng@fudan.edu.cn*).

The agents do not know, at $t = 0$ when the economy starts, whether the new technology is reliable or not. But if the technology is not reliable, it could be detected. Specifically, let $\nu \in [0, 1]$ be the measure (fraction) of the agents who are using the new technology in a given period – the adoption rate in that period. If the technology is *not* reliable, then with probability $1-P(\nu)$ the true state of the technology would be detected in the period. Assume $1 - P(\nu)$ is strictly increasing in ν. That is, the more producers are using the technology, the more likely an unreliable technology is detected in a given period.

The values of θ_0, θ_L, θ_H and the function $P(\nu)$ are all known to the producers, while the value of ν evolves endogenously. Assume the value of $P(\nu)$ is close to one for all values of ν – the technology, before being brought into the market, must have been tested many times, in the lab and by real users, each time producing the desired output θ_H. But θ_L is so bad that once the technology is found out to be not reliable – i.e., in case θ_L does occur – then it should be abandoned.

What happens at $t = 0$ and the periods after is for each individual producer to decide whether to adopt the new technology, and at what time. Those who have switched to the new technology could also make a decision to switch back. Switching to the new technology imposes a one time cost z on the producer, and the cost may differ between individual producers. Let the producers be distributed in z with a distribution density $g(z)$, where $z \in [0, \bar{z}]$, with $\bar{z} > 0$, and $g(z) \geq 0$ for all z. This distribution can be degenerate in which case the switching costs are a constant for all producers. Lastly, switching back from the new to the old technology imposes a common one time cost z_* on any producer.

At the beginning of $t = 0$ when the story starts, producers hold individual beliefs about whether or not the new technology is reliable. Let π ($\in [0, 1]$) denote an individual producer's belief in a given period of the probability with which the new technology is reliable. At $t = 0$ the producers are distributed over the interval $[0, 1]$, in their initial value of π. Let the density of the initial beliefs be denoted $f^0(\pi)$, $\pi \in [0, 1]$.

For technical reasons which become clear later, we denote the logit transformation of π as

$$\rho \equiv log\left(\frac{\pi}{1-\pi}\right), \tag{3.1}$$

and, to abuse notation a little bit, we let the distribution of the initial ρ be denoted also as $f^0(\rho)$, $\rho \in \mathbb{R}$. In the following, we use the word "belief" to refer to the producer's π or ρ.

As time unfolds and the new technology gets used by more producers, new information comes in, producers update their beliefs about the new technology, using a Bayesian updating rule. That is, suppose an individual producer starts with an initial belief ρ, and suppose the high output θ_H is produced by all users of the new technology in the current period, with ν being the current adoption rate. Then the producer's updated belief, ρ', is given by

$$\rho' = \rho - logP(\nu). \tag{3.2}$$

But if a low output θ_L is produced in any period, then all producers' new belief ρ' drops to $-\infty$ immediately. This implies that an unreliable new technology will be detected almost for sure in the long run. And once detected, the economy will convert permanently back to the old technology.

At the beginning of a period t, the economy is characterized by two aggregate states. The first is a measure of producers who are still with the old technology, $f_O^t(\rho)$, $\forall \rho \in \mathbb{R}$. Specifically, $f_O^t(\rho)$ is the measure of producers who are still with the old technology and hold the belief ρ at the beginning of period t. The second is a measure of producers who adopted the new technology in a prior period: $f_N^t(\rho)$, $\forall \rho \in \mathbb{R}$. Of course, $f_O^0 = f^0$ and $f_N^0(\rho) = 0$, $\forall \rho \in \mathbb{R}$.

It takes time to switch to the new technology from the old. In order to use the new technology in period t, the producer must make a decision to switch away from the old technology in period $t-1$. Thus the adoption rate in a period t is determined in period $t-1$ as

$$\nu_t = \int_{\mathbb{R}} f_N^t(\rho) d\rho. \tag{3.3}$$

(Note that this setup, though less intuitive, helps to simplify the value function.)

A producer who is still with the old technology then decides whether to switch to the new technology for the next period. A producer are using the new technology could also decide to switch back to the old. But in equilibrium, until the industry observes the first θ_L, all producers who had switched to the new technology will stay with the new technology.

Suppose a low output θ_L is observed in a period T. Then all producers using the new technology will cease production in $T+1$, and switch back to the old technology. This implies a value of V_* from period $T+2$ on, for all producers.

To close the model, we assume

$$u(\theta_0) \gg \lim_{\nu \to 0^+} P(\nu)u(\theta_H) + \big(1 - P(\nu)\big)u(\theta_L), \tag{3.4}$$

so that an unreliable technology, once detected, should be abandoned.

3.2 Equilibrium

At the start of a period, given that the new technology has not yet been detected as unreliable, individual producers are divided into two groups: those who are with the old technology, in state O, and those who are already with the new technology, in state N. Let $\nu \in [0, 1]$ denote the fraction of the producers that are already with the new technology, that is in state N. That is, ν is the "adoption rate" and we assume that its value is commonly observed by all producers. At the start of the period producers hold beliefs about whether the new technology is reliable or not. Let ρ denote an individual producer's belief that the new technology is reliable. This ρ may differ across individual producers. In a rational expectations equilibrium which we are about to define, agents are assumed to perfectly perceive the initial distribution of ρ across agents and how that distribution evolve over time. Specifically, let the rationally perceived measure of the beliefs of the state O producers be denoted f_O, and the measure of the beliefs of the producers in state N be denoted f_N.

Let the value of an individual producer in state $i(= O$ or $N)$ and with belief ρ at the start of a period be denoted $V_i(\rho; f_O, f_N)$. This agent must make a decision in the current period about whether to switch from his current technology state to the other technology and we let his choice be denoted $x_i(\rho; f_O, f_N)$, where $x_i = 1$ indicates the decision of choosing the new technology and $x_i = 0$ indicates the old. In a rational expectations equilibrium which we are describing, each agent takes as given his rationally perceived *equilibrium* distribution of the actions taken by all other producers in the current period, denoted $x_i^*(\rho; f_O, f_N)$.

Let us now describe the problem of an agent in state O, in the following Bellman equation:

$$\begin{aligned} V_O(\rho; f_O, f_N) &= \max_{x_O \in \{0,1\}} u(\theta_0) \\ &+ \beta \Big\{ (1 - x_O) \Big[\big(\pi + (1-\pi)P(\nu)\big) V_O(\rho'; f_O', f_N') + \underbrace{(1-\pi)\big(1 - P(\nu)\big)}_{\text{subjective prob of detection}} V_* \Big] \\ &+ x_O \Big[-\frac{z}{\beta} + \big(\pi + (1-\pi)P(\nu)\big) V_N(\rho'; f_O', f_N') + \underbrace{(1-\pi)\big(1 - P(\nu)\big)}_{\text{subjective prob of detection}} V_{**} \Big] \Big\} \end{aligned}$$

subject to

$$\pi = \frac{exp(\rho)}{1 + exp(\rho)}, \tag{3.5}$$

$$\nu = \int_{\mathbb{R}} f_N(\rho) d\rho, \tag{3.6}$$

$$\rho' = \rho - logP(\nu), \tag{3.7}$$

and for all $\hat{\rho} \in \mathbb{R}$,

$$f'_O(\hat{\rho}') = \sum_{i=O,N} \big[1 - \tilde{x}_i(\hat{\rho}; f_O, f_N)\big] f_i(\hat{\rho}); \tag{3.8}$$

$$f'_N(\hat{\rho}') = \sum_{i=O,N} \tilde{x}_i(\hat{\rho}; f_O, f_N) f_i(\hat{\rho}). \tag{3.9}$$

where $\hat{\rho}' \equiv \hat{\rho} - logP(\nu)$.

For a producer in state N, he takes as given $\tilde{x}_i(\rho; f_O, f_N)$, $i = O, N$ and solves:

$$\begin{aligned}
V_N(\rho; f_O, f_N) \quad &= \quad \underset{x_N \in \{0,1\}}{max} \big(\pi + (1-\pi)P(\nu)\big)u(\theta_H) + (1-\pi)\big(1 - P(\nu)\big)u(\theta_L) \\
&+ \quad \beta \Bigg\{ (1 - x_N)\Big[-\frac{z_*}{\beta} + \big(\pi + (1-\pi)P(\nu)\big)V_O(\rho'; f'_O, f'_N) + (1-\pi)\big(1 - P(\nu)\big)V_* \Big] \\
&+ \quad x_N \Big[\big(\pi + (1-\pi)P(\nu)\big)V_N(\rho'; f'_O, f'_N) + (1-\pi)\big(1 - P(\nu)\big)V_{**} \Big] \Bigg\}
\end{aligned}$$

subject to (3.5)-(3.9)

To make a distinction, we call i the individual agent's "technology state", and ρ, f_O, f_N the "belief states".

Notice that (3.7)-(3.9) describe the evolution of the belief states given that θ_L doesn't occur. The individual subjective probability of detecting a θ_L is given by $(1-\pi)\big(1 - P(\nu)\big)$. If θ_L is ever observed, all producers' belief about the reliability of the new technology will converge to $\pi = 0$ or $\rho \to -\infty$ and stay permanently there in an absorbing state. Given this, and given (3.4), we have two terminal values for the absorbing state:

$$V_* = \underset{\rho \to -\infty}{lim} V_O = \frac{u(\theta_0)}{1-\beta}; \tag{3.10}$$

$$V_{**} = \underset{\rho \to -\infty}{lim} V_N = -z_* + \beta V_*. \tag{3.11}$$

Definition 3. *A rational expectations equilibrium of the model consists of a set of value and decision functions for the individual agents,*

$$\big\{V_i(\rho; f_O, f_N),\, x_i^*(\rho; f_O, f_N) : i = O, N\big\},$$

and a law of motion for the economy's aggregate states $\{f_O, f_N\}$, given in equations (3.8)-(3.9), such that

(I) Given the evolution of the belief states, $\big\{V_i(\rho; f_O, f_N),\, x_i^*(\rho; f_O, f_N)\big\}$ solves agent i's Bellman equation, $i = O,\, N$.

(II) The evolution of the aggregate states ν, f_O and f_N, is consistent with the initial ν, f_O and f_N, the individual producer's optimal decisions and the Bayesian rule that the agents use for updating beliefs.

Following from (II) of the definition, if a low output (black swan event) is ever observed, all belief states degenerate to $\pi = 0$ ($\rho \to -\infty$). Otherwise, ρ evolves according to (3.7), f_i evolves according to (3.8) and (3.9).

3.3 Characterizing the Equilibrium

Assume $\underset{\nu \to 0^+}{lim} P(\nu) < 1$. That is, an unreliable new technology has a baseline probability of being detected, even if there are a very small measure of producers who have adopted it.

Proposition 5. *Suppose that the new technology is reliable. Suppose the initial distribution of beliefs doesn't have a mass point at $\pi = 0$. Then the equilibrium adoption rate will converge to 1 as time goes to infinity.*

But characterizing the dynamics on the equilibrium path is more challenging. In this section, we prove several properties of the equilibrium policy functions. With those properties, we can write an equivalent maximization problem, in which f_O, f_N are replaced by a cutoff belief $\bar{\rho}$ and the mean of the distribution of beliefs m_ρ.

Proposition 6. *In a rational expectations equilibrium of the model, $x_i(\rho; f_O, f_N)$ has the following cutoff property: For any given aggregate state (f_O, f_N), and $i = O, N$, there exists a cutoff belief $\bar{\rho}_i$ such that*

$$x_i(\rho; f_O, f_N) = \begin{cases} 1 & , \ \rho \geq \bar{\rho}_i \\ 0 & , \ \rho < \bar{\rho}_i \end{cases} .$$

Furthermore, we always have $\bar{\rho}_O \geq \bar{\rho}_N$.

The proof can be found in Appendix 3.5.4. The cutoff property of the policy functions is rather intuitive. It basically says that, given all other conditions equal, if a producer find it optimal to use the new technology, then all producers who hold higher beliefs - that the new technology is reliable - should also find it optimal to use the new technology. And $\bar{\rho}_O \geq \bar{\rho}_N$ says that, it is always easier to stick with the current technology, due to the switching costs.

Our goal is to find a tractable replacement of f_O, f_N on the equilibrium path. Knowing that the policy functions exhibit cutoff property, an immediate guess would be that the two measures of beliefs f_O, f_N are always separating the population distribution of beliefs f by some cutoff $\bar{\rho}$. The following proposition of cutoff rule verifies this guess.

Proposition 7. *Suppose the current aggregate state f_O, f_N are separating the population distribution of beliefs f by a cutoff $\bar{\rho}$. Following the policy functions, the next period f'_O, f'_N must also be separating f' by a cutoff $\bar{\rho}'$.*

The proof of the proposition is in Appendix 3.5.5. We start the economy with $f_O^0 = f^0$ and $f_N = 0$, which satisfy the cutoff rule. So we can replace the equilibrium dynamics of f_O, f_N by the equilibrium dynamics of $\bar{\rho}$ and f. Remember that from (3.2) we know $f^t(\rho)$ is always a shifting from $f^0(\rho)$. For a large set of distributions, this implies a change of mean, and we can replace f by its mean m_ρ. In conclusion, on the equilibrium path, we can use $\bar{\rho}$ and m_ρ as state variables instead.

One more thing to be noted before we go to the new maximization problem. The cutoff property of f_O, f_N could be violated in off-equilibrium situations. Think about the case where producers who hold belief ρ are in state N, producers who hold belief $\hat{\rho}$ are in state O, but $\bar{\rho}_O > \hat{\rho} > \rho > \bar{\rho}_N$. In the next period, this violation situation will continue, and presumably for longer periods. This is why we claim that the new maximization problem we are going to establish, is equivalent to the original one, only in the sense that they generate exactly the same equilibrium dynamics.

We rewrite the value functions and policy functions under three real-valued state variables: the individual belief ρ; The cutoff belief $\bar{\rho}$; And finally, the mean of the population distribution of beliefs m_ρ.

$$V_i(\rho; \bar{\rho}, m_\rho) : \ \mathbb{R}^3 \to \mathbb{R};$$

$$x_i(\rho; \bar{\rho}, m_\rho) : \ \mathbb{R}^3 \to \{0, 1\};$$

for $i = O, N$.

Because of the cutoff rule, we can also replace the rational perception of other producers' actions $\tilde{x}_i$, by a rational perception of the future cutoff $\tilde{\rho}'(\bar{\rho}, m_\rho)$. It can also be regarded as the aggregate policy function. We specify a functional form for

$$P(\nu) = \alpha^{\nu+c},$$

where $\alpha \in (0,1)$ and $c > 0$.[3] And write the following maximization problem.

Taking $\tilde{\rho}'(\bar{\rho}, m_\rho)$ as given, a producer with the old technology solves

$$\begin{aligned} V_O(\rho;\bar{\rho},m_\rho) \quad = \quad & \underset{x_O\in\{0,1\}}{max}\, u(\theta_0) + \beta\Bigg\{(1-x_O)\Bigg[\Big(\pi+(1-\pi)\alpha^{\nu+c}\Big)V_O(\rho';\bar{\rho}',m_\rho') + (1-\pi)(1-\alpha^{\nu+c})V_*\Bigg] \\ & +x_O\Bigg[-\frac{z}{\beta} + \Big(\pi+(1-\pi)\alpha^{\nu+c}\Big)V_N(\rho';\bar{\rho}',m_\rho') + (1-\pi)(1-\alpha^{\nu+c})V_{**}\Bigg]\Bigg\} \end{aligned}$$

s.t.

$$\pi = \frac{exp(\rho)}{1+exp(\rho)}; \tag{3.12}$$

$$\nu = \int_{\rho\geq\bar{\rho}} f(\rho|m_\rho)d\rho; \tag{3.13}$$

$$\rho' = \rho + (-log\alpha)(\nu+c); \tag{3.14}$$

$$\bar{\rho}' = \tilde{\rho}'(\bar{\rho}, m_\rho); \tag{3.15}$$

[3]See a discussion of why we pick this particular functional form in Appendix 3.5.2.

$$m'_\rho = m_\rho + (-log\alpha)(\nu + c); \tag{3.16}$$

And taking as given $\tilde{\rho}'(\bar{\rho}, m_\rho)$, a producer who is currently with the new technology solves

$$\begin{aligned} V_N(\rho; \bar{\rho}, m_\rho) &= \max_{x_N \in \{0,1\}} \big(\pi + (1-\pi)\alpha^{\nu+c}\big) u(\theta_H) + (1-\pi)(1-\alpha^{\nu+c}) u(\theta_L) \\ &\beta \Bigg\{ (1-x_N)\Bigg[-\frac{z^*}{\beta} + \big(\pi + (1-\pi)\alpha^{\nu+c}\big) V_O(\rho'; \bar{\rho}', m'_\rho) + (1-\pi)(1-\alpha^{\nu+c}) V_* \Bigg] \\ &+ x_N \Bigg[\big(\pi + (1-\pi)\alpha^{\nu+c}\big) V_N(\rho'; \bar{\rho}', m'_\rho) + (1-\pi)(1-\alpha^{\nu+c}) V_{**} \Bigg] \Bigg\} \end{aligned}$$

s.t.

$$(13) - (17)$$

$$V_* = \frac{u(\theta_0)}{1-\beta}; \tag{3.17}$$

$$V_{**} = -z_* + \beta V_*. \tag{3.18}$$

I claim that the maximization problem above is equivalent to the original problem, in the sense that they generate exactly the same equilibrium dynamics. At last, we shall describe the consistency condition for the rational expectation equilibrium. Define

$$\bar{\rho}_O(\bar{\rho}, m_\rho) = min\{\rho \in R \mid x_O(\rho; \bar{\rho}, m_\rho) = 1\},$$

$$\bar{\rho}_N(\bar{\rho}, m_\rho) = min\{\rho \in R \mid x_N(\rho; \bar{\rho}, m_\rho) = 1\},$$

then we can write the consistency condition as

$$\tilde{\rho}'(\bar{\rho}, m_\rho) = max\Big\{min\big\{\bar{\rho}, \bar{\rho}_O(\bar{\rho}, m_\rho)\big\}, \bar{\rho}_N(\bar{\rho}, m_\rho)\Big\} + (-log\alpha)(\nu + c). \tag{3.19}$$

where ν is derived from (3.13).

3.4 Numerical Simulations

We have solved the equilibrium numerically, you can find a detailed discussion of the algorithm in Appendix 3.5.6. We choose the following specification for the initial distribution of beliefs

$$f^0(\rho) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\bigg(- \frac{(\rho - \mu)^2}{2\sigma^2}\bigg),$$

that is, a normal distribution with mean μ and variance σ^2. In the benchmark parameterization, we use the following values:

Table 3.1: Benchmark Parameterization

Utility Levels	$u(\theta_0) = 0$, $u(\theta_H) = 1$, $u(\theta_L) = -15$
Probability	$\alpha = 0.7$, $c = 0.2$
Initial Distribution	$\mu = -2.2$, $\sigma = 1$
Grid Points	$N = 101$, $\pi_{min} = 10^{-7}$, $\pi_{max} = 1 - \pi_{min}$
Other	$\beta = 0.9$, $z = z^* = 1$

We choose the normal distribution in the ρ-space because it has a very good property: the initial mean of the distribution does not affect the equilibrium, only the initial variance does. The initial mean will come to play in the simulation part though. We have also tried with uniform distribution in the π-space, but find no qualitative difference.

Using the solved aggregate policy function $\tilde{\rho}'(\bar{\rho}, m_\rho)$, we have simulated the equilibrium path of adoption rate, starting from different initial means. The basic procedure for simulation is: we start from some given initial state $(\bar{\rho}^0, m_\rho^0)$, calculate the next period state from

$$\bar{\rho}^1 = \tilde{\rho}'(\bar{\rho}^0, m_\rho^0),$$

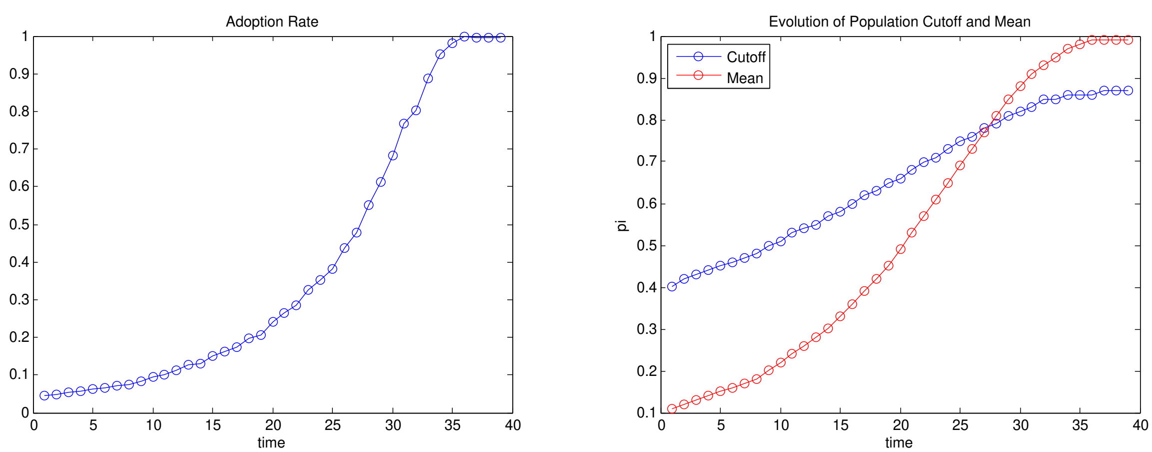
$$m_\rho^1 = m_\rho^0 + (-log\alpha)(\nu^0 + c),$$

and iterate for T many times. Notice that once $(\bar{\rho}^t, m_\rho^t)$ are given, the adoption rate ν^t can be calculated from

$$\nu^t = 1 - \Phi\left(\frac{\bar{\rho}^t - m_\rho^t}{\sigma}\right),$$

where Φ is the CDF of a standard normal distribution.

Figure 3.1: Adoption Rate and Evolution of Aggregate State



The adoption rate curve is clearly S-shaped. And we have this pattern for a large range of different initial means.

3.5 Appendix

3.5.1 The Extensive Form Problem

Though we are not going to solve it directly, let's write the extensive form maximization problem of a producer, and the corresponding rational expectations equilibrium(REE).

Starting with some initial belief π_0, technology $i = O, N$ and aggregate state $\left(f_O^0, f_N^0\right)$, the producer makes a history-contingent plan of all future choices of technology $\left\{x_t\right\}_{t=0}^{\infty}$. Let $x_t = 1$ indicates the choice of the new technology, and remember that the technology has a time-to-built, so x_t will become effective in $t+1$.

The history, which in general is a record of all past events the producer could observe, contains two components. The first is the history of all past outputs $\left\{\theta_s\right\}_{s=0}^{t-1}$. After knowing $\left\{\theta_s\right\}_{s=0}^{t-1}$, we can pin down the producer's past beliefs $\left\{\pi_s\right\}_{s=0}^{t}$ too. The second is the history of all producers' past choices of technology. Since we are interested in a symmetric equilibrium in which producers in the same state always make the same choice, this history is summarized by $\left\{f_O^s, f_N^s\right\}_{s=0}^{t}$. Define

$$h^t \equiv \left\{ \left\{\theta_s\right\}_{s=0}^{t-1} \; ; \; \left\{f_O^s, f_N^s\right\}_{s=0}^{t} \right\} \tag{3.20}$$

as a history, and H^t as the set of all possible histories, up to the beginning of period t. We can write the history-contingent plan of choices of technology as $\left\{x_t(h^t)\right\}_{t=0}^{\infty}$, the beliefs as $\left\{\pi_t(h^t)\right\}_{t=0}^{\infty}$, where

$$x_t(h^t): \; H^t \to \{0,1\};$$

$$\pi_t(h^t): \; H^t \to [0,1].$$

In an REE, the producer should be able to form a rational expectation of future dynamics of belief measures, which is history-contingent too. Define $\left\{\tilde{f}_O^{t+1}(\rho|h^t), \tilde{f}_N^{t+1}(\rho|h^t)\right\}_{t=0}^{\infty}$ as the producer's rational expectation of future dynamics of belief measures, where

$$\tilde{f}_i^{t+1}(\rho|h^t): \; H^t \to \mathcal{F}, \; i = O, N.$$

Now we can write the maximization problem, for an individual producer who holds initial belief π_0 or equivalently ρ_0, and in initial aggregate state (f_O^0, f_N^0). For notation consistency, we define $x_{-1} = 0$ $(= 1)$ if the producer starts at state O (N), and $h^0 = .$

The producer takes as given a rational expectation $\left\{\tilde{f}_O^{t+1}(\rho|h^t), \tilde{f}_N^{t+1}(\rho|h^t)\right\}_{t=0}^{\infty}$, and solves

$$\begin{aligned}
&\max_{\{x_t(h^t)\}_{t=0}^{\infty}} \mathbb{E}_0 \sum_{t=0}^{\infty} \beta^t \mathbb{E}_t \Bigg\{ -max\big\{x_t(h^t) - x_{t-1}, 0\big\} z - max\big\{x_{t-1} - x_t(h^t), 0\big\} z^* \\
&+\beta \Bigg[\big(1 - x_t(h^t)\big) u(\theta_0) + x_t(h^t) \Bigg[\Big(\pi_{t+1}(h^{t+1}) + \big(1 - \pi_{t+1}(h^{t+1})\big) P\big(\tilde{\nu}_{t+1}(h^t)\big) \Big) u(\theta_H) \\
&\qquad + \Big(1 - \pi_{t+1}(h^{t+1})\Big) \Big(1 - P\big(\tilde{\nu}_{t+1}(h^t)\big)\Big) u(\theta_L) \Bigg]\Bigg] \; \Bigg| \; h^t \Bigg\}
\end{aligned}$$

s.t. $\forall \; t$, conditional on $h^t \in H^t$

$$\tilde{\nu}_{t+1}(h^t) = \int_{\mathbb{R}} \tilde{f}_N^{t+1}(\rho|h^t) d\rho$$

$$\nu_t = \int_{\mathbb{R}} f_N^t(\rho) d\rho$$

$$\rho_{t+1}(h^{t+1}) = \begin{cases} \rho_t - logP(\nu_t) & , \; if \; \theta_t = \theta_H \\ -\infty & , \; if \; \theta_t = \theta_L \end{cases} \tag{3.21}$$

$$\pi_{t+1}(h^{t+1}) = \frac{exp\big(\rho_{t+1}(h^{t+1})\big)}{exp\big(\rho_{t+1}(h^{t+1})\big)+1}$$

The first constraint says that, the producer forms a rational expectation of the next period adoption rate, which directly affects the current choice of x_t. The second constraint says that given h^t, which contains f_N^t, the producer knows for sure the current adoption rate, which affects belief updating. The third constraint is Bayesian updating, if a high output is observed, the producer's belief ρ moves up by $-logP(\nu_t)$. If a low output is observed, it drops to $-\infty$, and stays there. The last constraint is to transform belief back to a probability.

In equilibrium, the evolution of aggregate state must be consistent with individual's optimal choices, that is, conditional on h^t

$$f_O^{t+1}(\rho_{t+1}|h^t) = \sum_{i=O,N} \big(1-x_t(h^t)\big) f_i^t(\rho_t); \tag{3.22}$$

$$f_N^{t+1}(\rho_{t+1}|h^t) = \sum_{i=O,N} x_t(h^t) f_i^t(\rho_t); \tag{3.23}$$

where each ρ_{t+1} corresponds to a ρ_t based on (3.21). And at last, for the equilibrium to be a REE, we require that for $\forall\ t$, $\forall\ h^t \in H^t$

$$\tilde{f}_i^{t+1}(\rho|h^t) = f_i^{t+1}(\rho|h^t),\ \ \rho\in\mathbb{R};\ \ i=O,N.$$

3.5.2 The $P(\nu)$ Function

If we think of the experiment as a hypothesis test, $1-P(\nu)$ would be the power function, i.e. the probability of rejecting the null (the new technology is reliable) when the alternative (the new technology is unreliable) is true. We want it to satisfy

$$-P'(\nu) > 0,\ \ \lim_{\nu\to\infty} 1-P(\nu) = 1.$$

The power of the test should be increasing with its size, and approaches a detection for sure in the limit.

An extra property we want this power function to have is that, if we run two experiments, one is with two stages of size ν_1 and ν_2, the other is with only one stage of size $\nu_1 + \nu_2$. Then the final belief we get should be the same. In other words, the size of the experiment should be accumulative.

One can prove that $P(\nu) = \alpha^\nu$ is the only continuous function satisfying the property. To see this, in the first experiment, we update belief in the following manner,

$$\begin{aligned}
\rho' &= \rho - logP(\nu_1), \\
\rho'' &= \rho' - logP(\nu_2) \\
&= \rho - logP(\nu_1) - logP(\nu_2),
\end{aligned}$$

while in the second experiment, we have

$$\hat{\rho}'' = \rho - logP(\nu_1 + \nu_2).$$

To such that $\rho'' = \hat{\rho}''$, we need

$$P(\nu_1 + \nu_2) = P(\nu_1)P(\nu_2), \ \forall \ \nu_1, \nu_2.$$

This uniquely pins down the functional form for $P(\nu) = \alpha^\nu$, as long as we assume that $P(\nu)$ is continuous. The detection probability is $1 - \alpha^\nu$ then.

Interestingly, this form is an analogue of the detection probability in the canonical Bayesian learning. Suppose we have the following signal structure, if the new technology is reliable (unreliable), it produces the bad outcome with probability 0 $(1 - a)$. Now we run $n \in \mathbb{Z}$ many i.i.d. experiments, one bad outcome is enough for us to conclude that the new technology is unreliable, so the detection probability is $1 - a^n$.

However, there are technical difficulties going directly from finitely many producers, to a continuum of them. Denote N as the total number of producers in the finite case, our ν resembles $\lim\limits_{n,N\to\infty} n/N$, and we can rewrite $1-a^n$ as $1-(a^N)^{n/N}$. What we want is $\lim\limits_{N\to\infty} a^N = \alpha$, but there is no fixed a making it happen. We need a to go up to 1 as N goes to infinity. In other words, we need the informativeness of the individual signal, which is determined by the difference between 0 and $1 - a$ in our case, to diminish when the number of signals goes up. In the literature, it is common to directly assume $P(\nu)$ with a continuum of producers, rather than to derive it as the limit of some finite cases.

At last, for technical reasons involving the definition of equilibrium, we can't allow $\lim\limits_{\nu\to 0^+} P(\nu) = 1$. Thus we choose $P(\nu) = \alpha^{\nu+c}$, where $c > 0$ is a constant.[4] The functional form doesn't follow the accumulative property exactly, but it is the closest we can get.

3.5.3 An Alternative Signal Distribution

Other signal forms could also work. For example, suppose the per capita output

$$\bar{Y} \sim N\big(\theta_H, \frac{\sigma^2}{\nu + c}\big)$$

when the new technology is reliable. And

$$\bar{Y} \sim N\big(\theta_L, \frac{\sigma^2}{\nu + c}\big)$$

when the new technology is unreliable.

We would have

[4] It is equivalent to $P(\nu) = c\alpha^\nu$, where $c \in (0, 1)$.

$$\rho' = \rho + \frac{(\nu + c)\left(\theta_L^2 - \theta_H^2\right)}{2\sigma^2} + \frac{(\nu + c)\left(\theta_H - \theta_L\right)}{\sigma^2}\bar{Y}.$$

It is still nice and clean, but in this case the belief could move up and down repeatedly, since there is no absorbing state. We leave it for future research.

3.5.4 Proof of Proposition 6

The proof consists of two steps. First we are going to prove that $V_N(\rho; f_O, f_N) - V_O(\rho; f_O, f_N)$ is non-decreasing in ρ. Then we will show that $\bar{\rho}_i,\ i = O, N$ exist, and $\bar{\rho}_O \geq \bar{\rho}_N$.

Step1: For any given (f_O, f_N), the difference term $\Delta V(\rho; f_O, f_N) \equiv V_N(\rho; f_O, f_N) - V_O(\rho; f_O, f_N)$ is non-decreasing in $\rho \in \mathbb{R}$.

To start with, notice that both $V_i(\rho; f_O, f_N),\ i = O, N$ are continuous in $\rho \in \mathbb{R}$, and differentiable in $\rho \in \mathbb{R}$ except for some measure 0 set of points.

In a state $i = O$ producer's maximization problem, the producer essentially compares

$$EV_O \equiv \big(\pi + (1-\pi)P(\nu)\big)V_O(\rho'; f'_O, f'_N) + (1-\pi)\big(1 - P(\nu)\big)V_*;$$
$$EV_N \equiv -\frac{z}{\beta} + \big(\pi + (1-\pi)P(\nu)\big)V_N(\rho'; f'_O, f'_N) + (1-\pi)\big(1 - P(\nu)\big)V_{**}.$$

Since we know that

$$\lim_{\rho \to \infty} V_O(\rho; f_O, f_N) = u(\theta_0) - z + \frac{\beta u(\theta_H)}{1-\beta};$$
$$\lim_{\rho \to \infty} V_N(\rho; f_O, f_N) = \frac{u(\theta_H)}{1-\beta};$$

and

$$u(\theta_H) - u(\theta_0) + z - \frac{z}{\beta} > 0.$$

We can conclude that $EV_N > EV_O$, as $\rho \to \infty$. In other words, there exists a $\hat{\rho}_O$ such that for $\forall\ \rho \geq \hat{\rho}_O$, $x_O(\rho; f_O, f_N) = 1$. Similarly, we can argue this for $x_N(\rho; f_O, f_N)$ and a $\hat{\rho}_N$.

Define $\hat{\rho} = max\big\{\hat{\rho}_O, \hat{\rho}_N\big\}$, then for $\forall\ \rho \geq \hat{\rho}$, we have $x_O(\rho; f_O, f_N) = x_N(\rho; f_O, f_N) = 1$ and

$$\Delta V(\rho; f_O, f_N) = \big(\pi + (1-\pi)P(\nu)\big)u(\theta_H) + (1-\pi)\big(1 - P(\nu)\big)u(\theta_L) - u(\theta_0) + z,$$

which is non-decreasing in π (and ρ). Thus $\Delta V(\rho; f_O, f_N)$ is non-decreasing in ρ on $[\hat{\rho}, \infty)$.

As for $\rho \in (-\infty, \hat{\rho})$, because of the differentiability, we can discuss the slope of $\Delta V(\rho; f_O, f_N)$ at any point, except for a measure 0 set of points. For any particular ρ, there are three cases to consider.

Case1: $x_O(\rho; f_O, f_N) = x_N(\rho; f_O, f_N) = 0$. We have

$$\Delta V(\rho; f_O, f_N) = \big(\pi + (1-\pi)P(\nu)\big)u(\theta_H) + (1-\pi)\big(1 - P(\nu)\big)u(\theta_L) - u(\theta_0) - z_*.$$

Obviously it is non-decreasing at such ρ.

Case2: $x_O(\rho; f_O, f_N) = x_N(\rho; f_O, f_N) = 1$. Same as those $\rho \in [\hat{\rho}, \infty)$.

Case3: $x_O(\rho; f_O, f_N) = 0,\ x_N(\rho; f_O, f_N) = 1$. We have

$$\begin{aligned}\Delta V(\rho; f_O, f_N) &= \big(\pi + (1-\pi)P(\nu)\big)u(\theta_H) + (1-\pi)\big(1 - P(\nu)\big)u(\theta_L) - u(\theta_0) \\ &+ \big(\pi + (1-\pi)P(\nu)\big)\Delta V(\rho'; f'_O, f'_N) + (1-\pi)\big(1 - P(\nu)\big)(V_{**} - V_*),\end{aligned}$$

where $\rho' = \rho - logP(\nu)$. As long as we can prove that $\Delta V(\rho'; f'_O, f'_N)$ is non-decreasing at ρ', $\Delta V(\rho; f_O, f_N)$ would be non-decreasing at ρ. Also notice that ρ' is greater than ρ by $-logP(\nu) > 0$. We can apply the same logic to $\Delta V(\rho'; f'_O, f'_N)$ to get a ρ'', and iterate this process until we reach some $\rho^n \in [\hat{\rho}, \infty)$. Then, use backward induction, we can show that $\Delta V(\rho; f_O, f_N)$ is non-decreasing at this ρ.

Note that $x_O(\rho; f_O, f_N) = 1,\ x_N(\rho; f_O, f_N) = 0$ is trivially impossible. So by its continuity, we can conclude that $\Delta V(\rho; f_O, f_N)$ is non-decreasing in $\rho \in \mathbb{R}$.

Step2: For any given (f_O, f_N), the cutoff beliefs $\bar{\rho}_i,\ i = O, N$ exist, and $\bar{\rho}_O \geq \bar{\rho}_N$.

Let's prove it for the state $i = O$ producers first. To show that $\bar{\rho}_O$ exists, it suffices to show that suppose $x_O(\rho_1; f_O, f_N) = 1$ and $\rho_2 > \rho_1$, then $x_O(\rho_2; f_O, f_N) = 1$ too.

$x_O(\rho_1; f_O, f_N) = 1$ implies that

$$\Delta V(\rho'_1; f'_O, f'_N) > \frac{(1-\pi_1)\big(1-P(\nu)\big)\big(u(\theta_0)+z_*\big)+z/\beta}{\pi_1+(1-\pi_1)P(\nu)}.$$

Now with a larger ρ_2, the LHS changes to $\Delta V(\rho'_2; f'_O, f'_N)$ where $\rho'_2 > \rho'_1$. Since we have proved that $\Delta V(\rho; f_O, f_N)$ is non-decreasing in $\rho \in \mathbb{R}$, we have $\Delta V(\rho'_2; f'_O, f'_N) \geq \Delta V(\rho'_1; f'_O, f'_N)$.

The RHS changes to

$$\frac{(1-\pi_2)\big(1-P(\nu)\big)\big(u(\theta_0)+z_*\big)+z/\beta}{\pi_2+(1-\pi_2)P(\nu)},$$

where $\pi_2 > \pi_1$, and obviously it is smaller.

Thus we know

$$\Delta V(\rho'_2; f'_O, f'_N) > \frac{(1-\pi_2)\big(1-P(\nu)\big)\big(u(\theta_0)+z_*\big)+z/\beta}{\pi_2+(1-\pi_1)P(\nu)}$$

must hold, and $x_O(\rho_2; f_O, f_N) = 1$. That concludes our proof for the existence of $\bar{\rho}_O$.

Proof for $\bar{\rho}_N$ follows the same logic, so we omit it here. What's left is to show that $\bar{\rho}_O \geq \bar{\rho}_N$. The proof is straightforward, remember that a state $i = O$ producer compares EV_O and EV_N. If we define the corresponding terms for a state $i = N$ producer, $\hat{EV}_O$ and $\hat{EV}_N$, we can easily check that

$$EV_O > \hat{EV}_O \ ; \ EV_N < \hat{EV}_N.$$

This implies that if a state $i = O$ producer finds it optimal to adopt the new technology, then a state $i = N$ producer, who is otherwise holding the same belief and in a same aggregate state, must also find it optimal to keep using the new technology. This immediately gives us $\bar{\rho}_O \geq \bar{\rho}_N$. QED

3.5.5 Proof of Proposition 7

By Theorem 1, we know the existence of $\bar{\rho}_O$ and $\bar{\rho}_N$. Based on the values of them and the current cutoff $\bar{\rho}$, there are three cases possible.

Case1: $\bar{\rho} > \bar{\rho}_O \geq \bar{\rho}_N$. Remember that producers with belief $\rho \in (-\infty, \bar{\rho})$ are currently in state $i = O$. Since $\bar{\rho} > \bar{\rho}_O$, those state $i = O$ producers with belief $\rho \in [\bar{\rho}_O, \bar{\rho})$ will switch to the new technology. And the next period cutoff is

$$\bar{\rho}' = \bar{\rho}_O - logP(\nu),$$

the adoption rate increases.

Case2: $\bar{\rho}_O \geq \bar{\rho} \geq \bar{\rho}_N$. In this case, neither state $i = O$ or state $i = N$ producer would want to change their current technology, and the next period cutoff is

$$\bar{\rho}' = \bar{\rho} - logP(\nu),$$

the adoption rate doesn't change.

Case3: $\bar{\rho}_O \geq \bar{\rho}_N > \bar{\rho}$. In this case, state $i = N$ producers with belief $\rho \in [\bar{\rho}, \bar{\rho}_N)$ will switch back to the old technology. And the next period cutoff is

$$\bar{\rho}' = \bar{\rho}_N - logP(\nu),$$

the adoption rate decreases.

We are confident that the Case3 won't happen on equilibrium path. But nevertheless, Proposition 2 holds, on equilibrium or off equilibrium path. QED

3.5.6 Algorithm for Numerical Solution

The model is solved numerically using a hybrid of value function iteration and policy function iteration technics. The algorithm contains two loops, the inner loop conducts value function iteration. While fixing the current guess of the aggregate policy function $\tilde{\rho}'(\bar{\rho}, m_\rho)$, the inner loop iterates on the two value functions $V_O(\rho; \bar{\rho}, m_\rho)$, $V_N(\rho; \bar{\rho}, m_\rho)$, until they both converge.

The outer loop conducts policy function iteration. It uses the converged value functions to generate a new aggregate policy function. Remember that those value functions are derived from the inner loop while fixing the current guess of the aggregate policy function. If the new aggregate policy function and the current guess of aggregate policy function are close enough, we are done.

Here are some more details. First, the state space is discretized. Originally, these three state variables ρ, $\bar{\rho}$, $m_\rho \in \mathbb{R}$, while their counterparts π, $\bar{\pi}$, $m_\pi \in [0, 1]$. We take an even grid in the π-space, that is, N grid points between some $[\pi_{min}, \pi_{max}]$. Then we use the logit transformation to get a corresponding uneven grid in the ρ-space. Thus in the loop, $V_O(\rho; \bar{\rho}, m_\rho)$ and $V_N(\rho; \bar{\rho}, m_\rho)$ are stored as $N \times N \times N$ matrices, while $\tilde{\rho}'(\bar{\rho}, m_\rho)$ is stored as a $N \times N$ matrix.

Second, to avoid coarseness from discretization, we allow the $\tilde{\rho}'(\bar{\rho}, m_\rho)$ matrix to take values in $\mathbb{R}$ (or in $[0,1]$, depending on its format). And when we need to map the value back on the grid, we use linear interpolation. For example, suppose $\tilde{\rho}' \in \big[\rho_{\text{grid}}(n), \rho_{\text{grid}}(n+1)\big]$, we have

$$V_i(n_\rho; \tilde{\rho}', n_{m_\rho}) = \frac{\rho_{\text{grid}}(n+1) - \tilde{\rho}'}{\rho_{\text{grid}}(n+1) - \rho_{\text{grid}}(n)} V_i(n_\rho; n, n_{m_\rho}) + \frac{\tilde{\rho}' - \rho_{\text{grid}}(n)}{\rho_{\text{grid}}(n+1) - \rho_{\text{grid}}(n)} V_i(n_\rho; n+1, n_{m_\rho}).$$

or use π_{grid} instead, when its format is a $\tilde{\pi}'$.

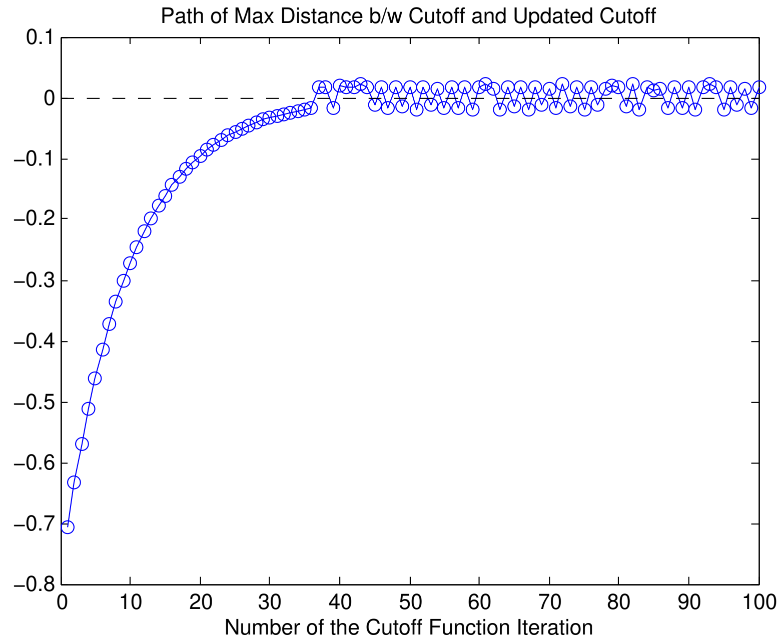
As in any typical iteration algorithm, we calculate the differences between the current guess of function and the newly derived function. When the distance, defined as the maximum absolute value of the differences, is small enough, we have ourselves a fixed point and that gives us the numerical solution. Since they are all real-valued functions, the convergence criteria for the value functions and the aggregate policy function are certain thresholds. The threshold for the value function iteration is set to be 10^{-7}. And the convergence in the value function iteration is universal, cause it is a contraction mapping.

The convergence in the policy function iteration is complicated though, we don't get universal convergence. At most of the grid points, convergence is achieved within a few rounds of iteration. But at some of the grid points, we find the aggregate policy function oscillate. For example, with $N = 101$ and 100 rounds of iteration, we end up with large differences[5] at 36 out of 101^2 grid points. And along the way, the set of grid points at which the differences are large becomes stable. Here is a graph of the maximum difference[6] during the 100 rounds of iteration.

[5] A difference is regarded as large if its absolute value is no less than 10^{-3}.

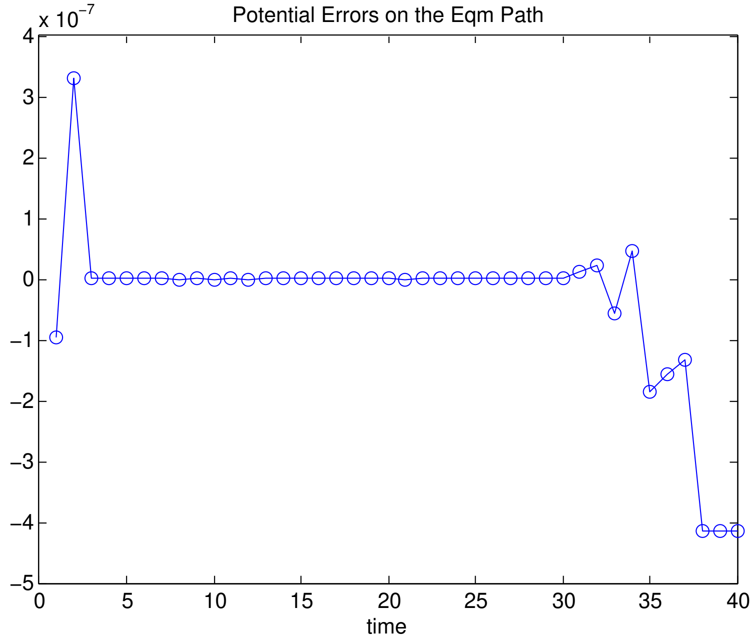
[6] In terms of the absolute value of course. In this numerical exercise, they are all achieved at the same grid point, except for the first few rounds of iteration.

Figure 3.2: Maximum Difference during Iterations



Those grid points at which the aggregate policy function oscillates, though being a very small portion of the total grid points, could jeopardize our simulation if they were on the equilibrium path. So the next task is to show that they are never visited, unless we start the economy there. We run several simulations, by varying the mean of the initial distribution from $\rho = -2.2$ ($\pi = 0.1$) to $\rho = 0$ ($\pi = 0.5$). In each simulation, we record the differences along the equilibrium path of state. It turns out that those differences never exceed a scale of 10^{-6}. Here is a graph of them when the initial mean is $\rho = -2.2$ ($\pi = 0.1$).

Figure 3.3: Errors along the Equilibrium Path



We conclude that the simulation results are valid, though we don't get universal convergence with the aggregate policy function. We could also reach such conclusion by switching to a more lenient measure of distance, for example, a weighted average $\frac{1}{N}\sqrt{\sum_{i,j}\text{difference}_{i,j}^2}$.

References

Chapter 1 References

[1] Acemoglu, Daron, Philippe Aghion and Fabrizio Zilibotti (2003): "Distance to Frontier, Selection, and Economic Growth." *Journal of the European Economic Association*, Vol. 4(1): 37-74.

[2] Aghion, Philippe and Peter Howitt (1992): "A Model of Growth through Creative Destruction." *Econometrica*, Vol. 60(2): 323-351.

[3] Baron, David P. (1982): "A Model of the Demand for Investment Banking Advising and Distribution Services for New Issues." *Journal of Finance*, Vol. 37(4): 955-976.

[4] Benston, George J. and Clifford W. Smith (1976): "A Transactions Cost Approach to the Theory of Financial Intermediation." *Journal of Finance*, Vol. 31(2): 215-31.

[5] Bernanke, Ben and Mark Gertler (1989): "Agency Costs, Net Worth, and Business Fluctuations." *American Economic Review*, Vol. 79(1): 14-31.

[6] Bernanke, Ben, Mark Gertler and Simon Gilchrist (1996): "The Financial Accelerator and the Flight to Quality." *The Review of Economics and Statistics*, Vol. 78(1): 1-15.

[7] Bernanke, Ben, Mark Gertler and Simon Gilchrist (1999): "The Financial Accelerator

in a Quantitative Business Cycle Framework." *Handbook of Macroeconomics*, Vol. 1(c): 1341-1393.

[8] Blackburn, Keith and Victor T. Y. Hung (1998): "A Theory of Growth, Financial Development and Trade." *Economica*, Vol. 65: 107-124.

[9] Buera, Franciso J., Joseph P. Kaboski, and Yongseok Shin (2011): "Finance and Development: A Tale of Two Sectors." *American Economic Review*, Vol. 101(5): 1964-2002.

[10] Campbell, Tim S. and William A. Kracaw (1980): "Information Production, Market Signalling, and the Theory of Financial Intermediation." *Journal of Finance*, Vol. 35(4): 863-882.

[11] Chemmanur Thomas J. and Paolo Fulghieri (1994): "Investment Bank Reputation, Information Production, and Financial Intermediation." *Journal of Finance*, Vol. 49(1): 57-79.

[12] Chemmanur Thomas J. and Paolo Fulghieri (1999): "A Theory of the Going-Public Decision." *Review of Financial Studies*, Vol. 12(2): 249-79.

[13] Chen Hsuan-Chi and Jay R. Ritter (2000): "The Seven Percent Solution." *Journal of Finance*, Vol. 55(3): 1105-1131.

[14] Cooley, Thomas F. and Edward C. Prescott (1995): "Economic Growth and Business Cycles." *Frontiers of Business Cycle Research*. Princeton University Press.

[15] Diamond, Douglas W. (1984): "Financial Intermediation and Delegated Monitoring." *Review of Economic Studies*, Vol. 51(3): 393-414.

[16] Diamond, Douglas W. and Philip H. Dybvig (1983): "Bank Runs, Deposit Insurance, and Liquidity." *Journal of Political Economy*, Vol. 91(3): 401-419.

[17] Galetovic, Alexander (1996): "Specialization, Intermediation, and Growth." *Journal of Monetary Economics*, Vol. 38(3): 549-559.

[18] Goldsmith, Raymond W. (1969): *Financial Structure and Development*. New Haven: Yale University Press.

[19] Greenwald, Bruce, Joseph Stiglitz and Andrew Weiss (1984): "Informational Imperfections in the Capital Market and Macroeconomic Fluctuations." *American Economic Review*, Vol. 74(2): 194-199.

[20] Greenwood, Jeremy, Juan M. Sanchez and Cheng Wang (2010): "Financing Development: The Role of Information Costs." *American Economic Review*, Vol. 100(4): 1875-1891.

[21] Greenwood, Jeremy, Juan M. Sanchez and Cheng Wang (2012): "Quantifying the Impact of Financial Development on Economic Development." *Review of Economic Dynamics*, Vol. 16(1): 194–215.

[22] Greenwood, Jeremy, Zvi Hercowitz and Gregory W. Huffman (1988): "Investment, Capacity Utilization, and the Real Business Cycle." *American Economic Review*, Vol. 78(3): 402-417.

[23] Gurley, John G. and Edward Stone Shaw (1962): *Money in a Theory of Finance*. Brookings Institution.

[24] Justiniano, Alejandro, Giorgio E. Primiceri and Andrea Tambalotti (2010): "Investment Shocks and Business Cycles." *Journal of Monetary Economics*, Vol. 57(2): 132–145.

[25] Justiniano, Alejandro, Giorgio E. Primiceri and Andrea Tambalotti (2010): "Investment Shocks and the Relative Price of Investment." *Review of Economic Dynamics*, Vol. 14(1): 101-121.

[26] King, Robert G. and Ross Levine (1993b): "Finance, Entrepreneurship and Growth: Theory and Evidence." *Journal of Monetary Economics*, Vol. 32(3): 513-542.

[27] Kiyotaki, Nobuhiro and John Moore (1997): "Credit Cycles." *Journal of Political Economy*, Vol. 105(2): 211-248.

[28] Leland, Hayne E. and David H. Pyle (1977): "Informational Asymmetries, Financial Structure, and Financial Intermediation." *Journal of Finance*, Vol. 32(2): 371-87.

[29] McGrattan, Ellen R. (2017): "Intangible Capital and Measured Productivity." *Staff Report 545*, Federal Reserve Bank of Minneapolis.

[30] Morales, Maria F. (2003): "Financial Intermediation In A Model Of Growth Through Creative Destruction." *Macroeconomic Dynamics*, Vol. 7(3): 363-393.

[31] Peter Alessandra (2019): "Owning Up: Closely Held Firms and Wealth Inequality." *Working Paper*.

[32] Ritter, Jay R. (2003): "Differences between European and American IPO Markets." *European Financial Management*, Vol. 9(4): 413-543.

[33] Veldkamp, Laura and Justin Wolfers (2007): "Aggregate Shocks or Aggregate Information? Costly Information and Business Cycle Comovement." *Journal of Monetary Economics*, Vol. 54(1): 37-55.

[34] Williamson, Stephen D. (1987): "Financial Intermediation, Business Failures, and Real Business Cycles." *Journal of Political Economy*, Vol. 95(6): 1196-1216.

Chapter 2 References

[35] Barkai, Simchar (2017): “Declining Labor and Capital Shares.” *working paper.*

[36] Boldrin, Michele and Lijun Zhu (2019): “Technical Progress and Movements in the Labor Share.” *working paper.*

[37] David, Joel (2017): “The Aggregate Implications of Mergers and Acquisitions.” *working paper.*

[38] De Loecker, Jan, and Frederic Warzynski (2012): “Markups and Firm-Level Export Status.” *American Economic Review*, Vol. 102: 2437-2471.

[39] De Loecker Jan, and Jan Eeckhout (2017): “The Rise of Market Power and the Macroeconomic Implications.” *working paper.*

[40] Edmond, Chris, Virgiliu Midrigan, and Daniel Yi Xu (2018): “How Costly are Markups.” *working paper.*

[41] Fox, Eleanor M. (2012): *Cases and Materials on U.S. Antitrust in Global Context.* 3rd edition, published by Thomson Reuters.

[42] Grullon, Gustavo, Yelena Larkin, and Roni Michaely (2016): “Are US Industries Becoming More Concentrated?” *working paper.*

[43] German, Gutierrez, and Thomas Philippon (2017): “Declining Competition and Investment in the US.” *working paper.*

[44] German, Gutierrez, and Thomas Philippon (2018): “How EU Markets Became More Competitive than US Markets: A Study of Institutional Drift.” *working paper.*

[45] Kimball, Miles S. (1995): “The Quantitative Analytics of the Basic Neomonetarist Model.” *Journal of Money, Credit, and Banking*, Vol. 27: 1241-1277.

[46] Klenow, Peter J., and Jonathan L. Willis (2016): "Real Rigidities and Nominal Price Changes." *Economica*, Vol. 83: 443-472.

[47] Levinsohn, James A., and Amil Petrin (2003): "Estimating Production Functions Using Inputs to Control for Unobservables." *Review of Economic Studies*, Vol. 70: 317-40.

[48] Mueller, Willard F. (1984): "Antitrust in the Reagan Administration." *Revue Francaise d'Etudes Armericaines*, 427-434.

[49] Olley, Steve G., and Ariel Pakes (1996): "The Dynamics of Productivity in the Telecommunications Equipment Industry." *Econometrica*, Vol. 64: 1263-1297.

Chapter 3 References

[50] Caselli, Francesco and Wilbur John Coleman II (2001): "Cross-country Technology Diffusion: The Case of Computers." *American Economic Review, Papers and Proceedings*, Vol. 91(2): 328-335.

[51] Caselli, Francesco and Daniel J. Wilson (2004): "Importing Technology." *Journal of Monetary Economics*, Vol. 51: 1-32.

[52] Chamley, Christophe P. (2003): *Rational Herds: Economic Models of Social Learning.* Cambridge University Press.

[53] Chen, Chaoran (2018): "Technology Adoption, Capital Deepening, and International Productivity Differences." *working paper.*

[54] Comin, Diego and Bart Hobijn (2004): "Cross-country Technology Adoption: Making Theories Face the Facts." *Journal of Monetary Economics*, Vol. 51: 39-83.

[55] Comin, Diego and Bart Hobijn (2010): "An Exploration of Technology Diffusion." *American Economic Review*, Vol. 100(6): 2031-59.

[56] Griliches, Zvi (1957): "Hybrid Corn: An Exploration in the Economics of Technical Change." *Econometrica*, Vol. 25(4): 501-522.

[57] Jovanovic, Boyan and Glenn MacDonald (1994): "Competitive Diffusion." *Journal of Political Economy*, Vol. 102(1): 24-52.

[58] Mansfield, Edwin (1961): "Technical Change and the Rate of Imitation." *Econometrica*, Vol. 29(4): 741-766.

[59] Manuelli, Rodolfo E. and Ananth Seshadri (2014): "Frictionless Technology Diffusion: The Case of Tractors." *American Economic Review*, Vol. 104(4): 1368-91.