Essays on Knowledge Spillovers

by

Jie Cai

B. E. Computer Science, Renmin University of China, 1999

M. A. Economics, Peking University, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF

Doctor of Philosophy

in

THE FACULTY OF GRADUATE STUDIES

(Economics)

The University of British Columbia

(Vancouver)

January 2011

Abstract

This thesis studies three issues involving knowledge diffusion across firms. The first chapter explains two data facts related to firm size distribution. First, it uses sector-specific inter-firm knowledge spillovers to explain sectoral differences in firm size heterogeneity. Greater inter-firm knowledge spillovers in a sector induce firms in that sector to invest relatively more in imitation. Greater imitation also causes faster catch-up by lagging firms and declining firm growth rate with firm size. Hence, the sectoral firm size distribution becomes more homogeneous in sectors with greater knowledge spillovers. Second, in a multi-sector version of this environment, I use inter-sector knowledge spillovers to explain the observed dependent *Pareto* size distributions in every subset of the economy. I test the model using patent citation data and find support for both its sectoral and aggregate predictions.

The second chapter rationalizes firms' motivation to build directed links with each other and formalizes the dynamic formation process that generates the observed network structure, including triple Power-law degree distributions, in the patent citation networks. Networks allow firms to become more specialized without losing customers, because having more firms in the market results not only in competitors but also in potential partner who redirect customers. Using firm citation panel data from the NBER Patent Citation Database, I estimate the model's parameters and simulate networks that exhibit similar structure features as corresponding real networks.

The third chapter documents a new empirical fact that larger firms update information faster than smaller firms in patent citation data and address its macroeconomic implications. In a model with size-dependent reaction time lag and *Pareto*

firm size distribution, the gradual spread of a firm-level technology shock generates a persistent and hump-shaped aggregate output growth rate. Greater information heterogeneity across firms de-synchronizes the co-movement among firms of different sizes, and hence causes a less volatile, smoother and longer aggregate business cycle. The model is well suited to explaining several timing relations of the business cycle. For example, productivity dispersion is pro-cyclical, the top firm's growth rate predicts future GDP growth, and investment leads hiring over the business cycle.

Table of Contents

List of Tables

List of Figures

Acknowledgments

I am heartily thankful to my supervisor, Amartya Lahiri, and my thesis committee members, Paul Beaudry and Patrick Francois, whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of the subject.

I would like to show my gratitude to the attendees at UBC Macroeconomics and DIET lunch seminars for your wonderful comments. I am indebted to many of my colleagues who supported and encouraged me during the PhD study.

Lastly, I offer my regards and blessings to all those who supported me in any respect during the completion of the thesis.

Dedication

To my parents and husband.

Chapter 1

Knowledge Spillovers and Firm Size Heterogeneity

1.1 Introduction

Increasingly firm- or establishment-level data show that firm size distributions within narrowly defined sectors and within the overall economy are widely dispersed and follow a *Pareto* distribution. Two important related questions are not well understood in the literature of firm growth dynamics. First, why is firm size heterogeneity[1] different across sectors? Second, why does *Pareto* firm size distribution exist in every subset of the economy; Moreover, why are firm size variables in different sectors dependent on each other?

This paper uses intra-sector and inter-sector knowledge spillovers, respectively, to answer these two questions. In a one-sector model, cross-sector differences in firm size heterogeneity can be attributed to sector-specific intra-sector knowledge diffusion efficiency. The one-sector model used in this paper extends the endogenous innovation model of Klette and Kortum (48) by giving firms the option to imitate. In sectors with more abundant knowledge spillovers, firms invest relatively more in imitation, as compared to innovation. Imitation then contributes a

[1] Using French, Chilean and U.S. firm-level data, Appendix A shows that sector-specific firm size heterogeneity is robust to different proxies of firm size, stable over time in the same country, and highly correlated across different countries in the same year.

greater share to the gross growth rate. Since equal opportunity to learn provide a stronger impetus to small firms, firm growth rate drops faster as the firm becomes larger. The sectoral firm size distribution is more homogeneous if small firms have more opportunities to catch up with the leaders. The model implications are confirmed using NBER (National Bureau of Economic Research) Patent Citation Data, which provides a measure of knowledge spillovers and the appropriate information for distinguishing the contributions by imitation and innovation to firm growth rate. The one-sector model suggests that optimal intellectual property rights depend on the trade-off between a higher imitation rate and a lower private return of knowledge.

A multi-sector model incorporates two additional facts that are absent in the one-sector model: firms develop products in multiple sectors, and inter-sector knowledge spillovers integrate the firm growth dynamics in all sectors. Firm growth dynamics in any subset of the economy are subject to a similar influence from all sectors, and the firm size distribution therefore converges universally to a *Pareto* distribution, not only within narrowly defined sectors but in the economy overall. Besides the implication on firm size distribution, the multi-sector model suggests one more channel than the one-sector model through which intellectual property rights protection promotes economic growth. Stronger intellectual property rights directly raise the private return of the intensive knowledge contributing sector more than other sectors, therefore attracting a larger share of research investment to the knowledge giver sectors. Indirectly, a better cross-sector R&D resource allocation that is justified by sectoral knowledge externality boosts economic growth rate.

The one-sector model in this paper extends that of Klette and Kortum (48) by allowing firms to create new goods by imitation, as well as by innovation. The difference between innovation and imitation is that innovation relies on a firm's private knowledge (measured by its current number of goods), while imitation relies on a sector's public knowledge (measured by the average firm size in the sector). Both types of R&D are subject to independent and identically distributed (i.i.d.) shocks, which are necessary to induce the *Pareto* firm size distribution. Sector-specific knowledge diffusion efficiency is given by a firm's productivity in utilizing private knowledge in innovation and public knowledge in imitation. When it is relatively more efficient to imitate than innovate, firms invest relatively more in imitation; as a result, the imitation rate contributes a greater share to the overall growth rate for the firm and the entire sector.

A scale-dependent firm's growth rate is the summation of the scale-independent innovation rate and the scale-dependent imitation rate. As specified in the Cobb-Douglas production function of new goods, output is proportional to the input. The innovation rate is independent of firm size, because the number of new goods generated from private knowledge is proportional to the firm's private knowledge capital. In contrast, the imitation rate decreases with firm size, because the number of imitated new goods is proportional to the public knowledge pool. When imitated output is divided by firm size, smaller firms obtain higher imitation rates. In sectors where imitation accounts for a greater share of the total growth rate, a firm's growth rates drop faster as the firms become larger.

A larger growth rate gap between smaller and larger firms causes faster firm size mean reversion and a more homogeneous firm size distribution. According to Kesten (44), when the innovation shock and imitation shock are i.i.d. across time and firm, firm size distribution within the sector converges to a *Pareto* distribution with scale parameter, while in the Klette and Kortum (48) environment without imitation, the firm size distribution converges to a logarithmic distribution. When the innovation risk follows a log-normal distribution, the closed form solution of the firm size heterogeneity measure, $1/\mu$, increases with the volatility of innovation shock and decreases with the imitation's contribution to the gross growth rate of the sector. Intuitively, the innovation shocks generate the firm size dispersion, while imitation shocks alleviate firms' size differences by allowing firms to learn and catch up.

The one-sector model used here has three testable implications. First, scale-dependent firm growth rate arises purely from the scale-dependent imitation rate. Specifically, a surviving firm's imitation rate drops as its size increases, while a firm's innovation rate is independent of firm size. Moreover, a firm's growth rate is more scale-dependent in sectors with more abundant knowledge spillovers and more homogeneous firm size. Second, firm size heterogeneity decreases with imitation's contribution to the gross growth rate, and increases with the volatility of innovation risk. Third, knowledge diffuses faster in sectors with a more homogeneous firm size distribution.

The challenge in testing the first two implications is to distinguish the imitation rate from the innovation rate in the total firm growth rate. This is done by differentiating between citations given to the citing firm's old patents (inside citations) and citations given to other firms' old patents (outside citations). A firm's quality adjusted growth rate of patent stock is split into innovation rate and imitation rate according to the ratio between inside citations and outside citations. Each outward citation is weighted by the importance of the cited firm[2] to control for the quality of knowledge spillovers embodied in each citation. When calculating a firm's quality adjusted growth rate, each new patent is also weighted by its number of quality adjusted citations received in the 10 years after application.

In the regression of innovation (imitation) rate on firm size, the coefficient is barely (always) statistically significantly different from zero among 42 sectors, as predicted by the second model implication. This result supports the idea that imitation rate declines with firm size, while innovation rate is independent of firm size. Moreover, the regression coefficient of imitation rate on firm size decreases with firm size heterogeneity, which confirms the first model prediction, that imitation rate is more scale-dependent in sectors with more homogeneous firm size distribution. With an estimated innovation rate and imitation rate for every firm within a sector, I can derive the sector-level imitation rate's share in gross growth rate and the variance of the log scale innovation rate. The second implication is also supported by the data: firm size heterogeneity is negatively related to imitation rate's share in gross growth rate and positively related to the volatility of innovation risk.

To test the third implication, I employ within-sector patent citations in NBER Patent Citation Data as a measure of intra-sector knowledge spillovers. I only include citations made by firms located in United States. The cross-firm knowledge diffusion speed and the percentage of cross-firm citations among all citations are negatively correlated with the firm size distribution heterogeneity. In the regressions, I control the geographic distance between the citing and the cited patent, size of the citing organization, cited organization and sector size. The knowledge

[2] I measure the importance of a citing firm f in sector s by its hub weight in the firm networks connected by cross-firm patent citations within sector s. The hub weight of each firm is calculated according to Kleinberg (45), which measures a node's ability to absorb information through the networks.

spillover speed is measured by citation time lag[3]. As a robustness check, the cross-sector difference in knowledge spillovers also holds in the citation data including all G7 countries.

Now, to turn to the second question: why does firm size distribution follow a *Pareto* distribution not only within narrowly defined sectors but also in the entire economy? According to Jessen and Mikosch (40), summation or pooling of independent *Pareto*-distributed variables induces a new *Pareto*-distributed variable. However, the scale parameter of the new *Pareto* distribution should be equal to the smallest scale parameter of the component distributions. In contrast, firm-level data (Figure 1.1 and Figure 1.2) show that the scale parameter of the size distribution for the whole economy is in the middle of the range of the component distributions' scale parameters. Therefore, some mechanism must make firm size and firm growth dynamics in different sectors dependent on each other.

The multi-sector model adds two important elements to the one-sector model. First, many firms develop products in multiple sectors. In the NBER Patent database, every organization, on average, applied for patents in 6.5 out of 42 patent categories; moreover, larger organizations cover more categories (see Table 1.5). Second, inter-sector knowledge spillovers are as important as within-sector knowledge diffusion. Inter-sector citations amount to 37% of total citations in the citation data. In the multi-sector model, when firms invent new products in a single sector, they can apply their private knowledge capital from all sectors. Also, a firm's growth in a single sector is affected by its previous knowledge capital in all sectors. As a result, a firm's overall size, which is the summation of its branches in all sectors, is influenced by its private knowledge capital in all sectors. Since a firm's growth dynamics in any subset of the economy follow a similar formula, firm size distribution converges to *Pareto* distribution universally in any subset of the economy.

The models demonstrate policy implications for optimal intellectual property rights protection (IPRP). In the one-sector model, the growth maximizing degree of protection depends on the trade-off between a higher imitation rate and a lower private return of knowledge. In the multi-sector model, cross-sector resource al-

[3] The citation lag, the time difference between the application time of the citing patent and that of the cited patent, indicates the time needed for the knowledge to travel between the citing patent inventor and the cited patent inventor.

location is another factor in the trade-off that promotes better IPRP. A strengthened IPRP increases the private return of major knowledge contributing sectors more than other sectors, hence encouraging more research investment to enter the knowledge giver sectors. Indirectly, economic growth benefits from a cross-sector resource allocation that agrees with each sector's knowledge contribution to the economy. Similarly, in the open economy setting of Cai and Li (11), the multi-sector model discovers a new channel through which trade cost harms growth: the distorted R&D resource allocation across-sectors, because trade cost reduces the private return of those sectors contributing intensive knowledge spillovers more than other sectors.

1.1.1 Literature Review

In the literature on firm growth dynamics, I have identified two strands in the theoretical debate about sources of firm size heterogeneity. The first strand, represented by the work of Lucas (51), Jovanovic (43), and Klette and Raknerud (49), emphasizes the impact of a manager's various talents in creating permanent differences in firm efficiency. The second strand, elaborated by Hopenhayn (35), Ericson and Pakes (25), Klepper (46), Klette and Kortum (48), Klepper and Thompson (47), and Luttmer (54), contends that firm size dispersion is caused by accumulated idiosyncratic shocks over a firm's life cycle. Seker (64) incorporates both of these strands and tries to distinguish the contribution of each.

In addition to exploring the origins of firm size heterogeneity, the literature on firm size dynamics tries to explain interesting stylized facts observed in firm-level data. First, the firm size distribution follows a *Pareto* distribution, both within individual sectors and in the entire economy. Meanwhile, the heterogeneity of firm size distribution varies across sectors (see Figure 1.1 and Figure 1.2), as documented by Axtell (4); Helpman, Melitz, and Yeaple (33); Rossi-Hansberg and Wright (63) (RW henceforth); and Luttmer (54). Second, a surviving firm's expected growth rate drops with firm size, or put differently, firm growth rate is scale-dependent[4].

Studies take various approaches when modeling scale-dependent surviving firm

[4] See Evans (26); Hall (30); Dunne, Roberts, and Samuelson (22); Sutton (68); Klette and Kortum (48); and Luttmer (54). Again, there are cross-sector differences in growth rate scale dependency, as demonstrated in RW.

growth rate. Cooley and Quadini (18), Cabral and Mata (10), Albuquerque and Hopenhayn (2) and Clementi and Hopenhayn (14) show that financial market friction can induce firm growth rate to decline with firm size. In Klette and Kortum (48), every firm has the same unconditional growth rate, while small firms have a higher growth rate conditional on survival, because they are less likely to survive than large firms. Selection is the key to having scale-dependent firm growth rate. Klepper and Thompson (47) use creation and de-construction of sub-markets to explain firm size dynamics. The size decrement due to sub-market de-construction is proportional to firm size, while the size increment due to the creation of emerging markets is independent of firm size. As the firm operates in more sub-markets, the expected proportional increase in size declines. RW shows that small firms grow faster because households want to accumulate industry-specific human capital more rapidly in small firms where the marginal return of human capital is still high. Luttmer (54) assumes that new firms enter with a high-quality blueprint, but that the blueprint's quality depreciates and becomes obsolete over time. Hence, firms choose to replicate their blueprints faster when they are smaller and their blueprints' quality is still high.

Only RW pays attention to the cross-sector differences in firm growth rate scale dependence and firm size heterogeneity in terms of sector-specific capital intensiveness. The firm growth rate drops faster in more capital intensive or less human capital intensive sectors because the marginal return to human capital decreases faster there. When small firms are more likely to catch up with large firms, firm size distribution becomes more homogeneous. Capital intensiveness can explain cross-sector differences in firm size heterogeneity for broad sector divisions (for instance, between education, construction and manufacturing.), but it fails to explain differences across a more refined division of manufacturing sectors. The patent citation data used in this paper primarily cover manufacturing and shows that examining knowledge spillovers efficiency is a more promising way to account for differences across manufacturing sectors.

Luttmer (53) and RW also consider the role of knowledge spillovers in shaping firm size distribution. In Luttmer (53), only entrants learn from incumbents, and as new entrants can learn more from incumbents, the firm size distribution is more homogeneous. In the learning-by-doing (LBD) extension of RW, the industrial total output enters the accumulation function of industry-specific human capital. The authors show that a larger externality in LBD leads to a faster mean reversion in human capital stock and a more homogeneous firm size distribution. The conclusion of the extension is that capital intensiveness and the LBD externality jointly determine firm size dispersion. In addition, the firm size distribution in their paper converges to a log-normal distribution, instead of a *Pareto* distribution. Unfortunately, unlike this paper, neither of the above two papers provides empirical evidence to support its theoretical predictions on knowledge spillovers.

This paper differs from Luttmer (53) in that it allows every firm, instead of just new entrants, to imitate. It is closer to RW's extension featuring LBD externality, but this paper uses a micro-founded approach, while RW uses a macroeconomic approach. In some sense, knowledge spillovers are one reason for the depreciation of a blueprint's quality in Luttmer (54). Expecting that others will 'steal' its blueprint in the future, the owner of a blueprint chooses to replicate it faster before others imitate it.

No prior research has studied the universal *Pareto* firm size distribution that is found across all subsets of the economy. In addition, the multi-sector model makes a distinct contribution by providing growth policy suggestions concerning intellectual property rights and trade policy through the cross-sector resource allocation channel.

The remainder of the paper is organized as follows. In section 2, the one-sector model shows that sector-specific intra-sector knowledge diffusion efficiency determines a firm's choice of endogenous innovation and imitation inputs, which in turn affects the firm size heterogeneity. In section 3, I test the implications of the one-sector model with NBER Patent Citation Data. The multi-sector model is presented in section 4, where I show that inter-sector knowledge spillovers integrate growth dynamics in all sectors and induce a *Pareto* size distribution in all subsets of the economy. Section 5 concludes.

1.2 One-Sector Model

1.2.1 Consumer

The representative consumer faces the following problem:

$$U = \max_{\{x_{i,t}\}} \sum_{t=0}^{\infty} \rho^t \left[\log\left(C_t\right)\right] \tag{1.1}$$

s. t.

$$P_t C_t + \int_0^{M_F} S_{f,t+1} V_{f,t} df = L + \int_0^{M_F} S_{f,t}\left(V_{f,t} + D_{f,t}\right) df$$

$$C_t = \left(\int_0^{I_t} x_{i,t}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}.$$

ρ is the time preference of the representative consumer; C_t is the consumption of final goods; the consumption and price of intermediate good i are x_{it} and $p_{i,t}$, respectively. P_t is the aggregate price index. The representative consumer inelastically supplies L hours of labor each period. Wage rate is normalized to 1. M_F is the total number of firms in the economy. $S_{f,t}$ is the equity share of firm f held by the consumer at time t. $V_{f,t}$ is the price of firm f equity at time t. $D_{f,t}$ is the dividend to firm f shareholders at time t.

There are I_t intermediate goods in the economy at time t. $\sigma > 1$ is the elasticity of substitution between intermediate goods. Consumer demand for intermediate goods is

$$x_{i,t} = C_t \left(\frac{P_t}{p_{it}}\right)^{\sigma}$$

$$P_t = \left(\int_0^{I_t} p_{it}^{1-\sigma} di\right)^{\frac{1}{1-\sigma}}$$

From the first order conditions for the consumer's problem, the equity price of firm f is given by

$$V_{f,t} = \rho \frac{C_t P_t}{C_{t+1} P_{t+1}} \left(V_{f,t+1} + D_{f,t+1}\right)$$

1.2.2 Firms

There is only one sector with M_F firms in the economy. M_F is a large number, so each firm is tiny relative to the economy. Firm f hires one unit of production labor to produce one unit of goods. The wage rate is the numeraire. According to Dixit and Stiglitz (20), the profit-maximizing price for every product is $\frac{\sigma}{\sigma-1}$. In the monopolistic competitive market, the profit from each product is $\frac{1}{\sigma}\frac{Y_t}{I_t}$. Firm f produces $z_{f,t}$ number of goods that it has invented by time t. The total number of goods in the economy is $I_t = \int_0^{M_F} z_{f,t} df$.

Firms grow by inventing new goods. Firm f invents the new goods through two types of R&D: innovation and imitation. Innovation uses firm's private knowledge capital and $N_{f,t}$ units of research hour. Here the size of firm f's private knowledge capital is measured by $z_{f,t}$ to represents firm f's experience in R&D. In contrast, imitation uses public knowledge capital $\bar{Z}_t$[5] and $M_{f,t}$ units of research hour.

The size of the public knowledge pool is measured by the average firm size $\bar{Z}_t$ in the industry. This assumption implies that learning is time-consuming and that no firm can afford to acquire all outside knowledge in one period. In other words, the inter-firm learning happens multiple-to-multiple instead of all-to-all. A number of pairs of firms are randomly matched to learn from each other. What one firm expects to learn from a random peer is the average firm size in the sector.

The assumption of proxy public knowledge by average firm size rather than total industry size, is supported by the firm citation network data. If the total industry size was the right measure of public knowledge pool size, we would observe that each firm is linked to every other firm, or a significant proportion of the total number, whereas network density, the average number of cited firms per firm is far smaller than the total number of firms. For example, in 1976 there were 5749 U.S. patenting firms, which on average cited patents from 2.81 firms; in 1996 there were 20,522 U.S. patenting firms, which on average cited patents from 11.02 firms. Although the network density increases with the total number of patenting firms, each firm still only learns from a limited number of peers. Additionally, this assumption ensures that in the general equilibrium, economic growth rate is a constant, which is independent of the total number of firms and population size.

[5] In Appendix A.4, firms use both private and public knowledge to imitate.

Imitation here does not refer to simple reverse engineering and replication; it means improvements and upgrades of other firms' products. This assumption reflects the fact that patent law does not acknowledge simple replication. In order to gain a new patent, a firm must upgrade existing patented goods to demonstrate enough originality and creativity. Firms' private knowledge diffuses to the public knowledge pool through many channels[6]. Firms may or may not voluntarily reveal their private knowledge to the public, but interactions between firms always generate a steady flow of knowledge from each firm's private pools to the public pool.

I borrow the Cobb-Douglas production function from Klette and Kortum (48)to describe the new goods production function. The expected number of new goods depends on the amount of hours invested in research and knowledge capital.

$$E\left(\Delta z_{f,t}^{N}\right) = A_N N_{f,t}^{\alpha} z_{f,t}^{1-\alpha}$$

$$E\left(\Delta z_{f,t}^{M}\right) = A_M M_{f,t}^{\alpha} \bar{Z}_{t}^{1-\alpha}$$

$\Delta z_{f,t}^{N}$ ($\Delta z_{f,t}^{M}$) is the number of new goods invented by innovation (imitation). $0 < \alpha < 1$ is the labor share in knowledge production.

The productivity of innovation (imitation) A_N (A_M) is sector-specific. This assumption captures the fact that technology is more standardized or codified in some sectors than others. Standardized industrial technology is easier to transplant across firms, while firm-specific technology is only suited for application within the inventing firm. Additionally, standardized industrial technology enables workers to change employers within the same sector. If knowledge capital is embodied in workers, a high labor turnover rate also helps to disseminate one firm's private knowledge to other firms.

Within a sector, the productivity in the two types of R&D A_N and A_M can be different, which implies that firms employ private and public knowledge at different costs. These costs include the searching cost of related existing knowledge, the

[6]In Duguet and MacGarvie (21), there are 12 channels listed: external R&D, cooperative R&D, patents and licenses, analysis of competition, experts, equipment acquisition, hiring employees, communication with suppliers, communication with customers, mergers and acquisitions, joint ventures and alliances, and personnel exchange.

reverse engineering cost of absorbing the existing knowledge, and the creation cost of adding novelty to the existing knowledge. Normally, firm borders block knowledge spillovers, and it is therefore more efficient to use private knowledge instead of public knowledge, i.e. $A_N > A_M$[7].

The Cobb-Douglas knowledge production function hinges on two assumptions. First, R&D research hours $N_{f,t}$ and $M_{f,t}$ have decreasing marginal productivity. Second, with the same amount of innovative research hours, $N_{f,t}$, larger firms invent more new goods due to greater R&D experience. Similarly, with the same amount of imitative research hours, $M_{f,t}$, firms with access to a deeper public knowledge pool $\bar{Z}_t$ create more new goods. When these two forces offset each other, the amount of research hours that each firm spends on innovation (imitation) is proportional to the size of private (public) knowledge pool $z_{f,t}$ ($\bar{Z}_t$).

Firm f chooses inputs in innovation $N_{f,t}$ and imitation $M_{f,t}$ to maximize its firm value $V(z_{f,t})$.

$$\max_{N_{f,t},\, M_{f,t}} V(z_{f,t}) = \frac{P_t C_t}{\sigma} \frac{z_{f,t}}{I_t} - \frac{N_{f,t} + M_{f,t}}{\bar{Z}_t} + \frac{\rho C_t P_t}{C_{t+1} P_{t+1}} E[V(z_{f,t+1})] \tag{1.2}$$

subject to

$$z_{f,t+1} = z_{f,t} + \Delta z^N_{f,t} + \Delta z^M_{f,t} \tag{1.3}$$

$$\frac{\Delta z^N_{f,t}}{z_{f,t}} = \frac{A_N N^{\alpha}_{f,t} z^{1-\alpha}_{f,t}}{z_{f,t}} + \varepsilon^n_{f,t} \tag{1.4}$$

$$\frac{\Delta z^M_{f,t}}{\bar{Z}_t} = \frac{A_M M^{\alpha}_{f,t} \bar{Z}^{1-\alpha}_t}{\bar{Z}_t} + \varepsilon^m_{f,t} \tag{1.5}$$

$\frac{z_{f,t}}{I_t}$ represents firm f's market share in terms of both number of goods and sales,

[7]On average, firms tend to use private knowledge more frequently and sooner than public knowledge. In the NBER Patent Citation Data, every organization on average owns 0.17% of the old patent stock in the industry, but the rate at which they cite their own old patents is disproportionately high at 11.1%. If the cited and citing patents are owned by the same organization, the average citation time lag is 5.86 years; otherwise the average time lag is 9.06 years. Citation lag is defined as the application year of citing patent minus the application year of cited patent, which indicates how long it takes the citing firm to acquire and make use of the knowledge embodied in the cited patent.

because each good is sold at the same amount. Firm f's investments in R&D decide the expected success rates of innovation and imitation, but the actual realization in (1.4) and (1.5) are subject to i.i.d. shocks $\varepsilon^n_{f,t}$ and $\varepsilon^m_{f,t}$ [8]. When firm f's manager chooses research inputs at the beginning of time t, she knows the distributions of $\varepsilon^n_{f,t}$ and $\varepsilon^m_{f,t}$ but not their actual realizations.

Firm f discounts future firm value at the same rate of consumer's $\frac{\rho C_t P_t}{C_{t+1}P_{t+1}}$. Labor productivity in R&D grows as fast as the public knowledge pool size $\bar{Z}_t$, because workers learn human capital from their employers; when workers turn over across firms, their productivity of R&D is equal to the average knowledge capital among firms [9]. Since each firm is tiny relative to the entire sector, firm f takes I_t, Y_t and P_t as given. I assume that firms receive full liquidation value in case of exit, so that their current innovation and imitation decisions are independent of exit risk in the future.

One educated guess for the firm value is a linear function of the form:

$$V(z_{f,t}) =$$

$v_t \frac{z_{f,t}}{I_t} + u_t - F$. (1.6)

v_t is the private marginal value of market share. I will show later that u_t represents the rent from public knowledge externality. F is the fixed entry cost to the market.

The first order conditions are:

$$N_{f,t} = \left(\frac{A_N \alpha v_{t+1} \rho'_t}{M_F}\right)^{\frac{1}{1-\alpha}} z_{f,t} \tag{1.7}$$

$$M_{f,t} = \left(\frac{A_M \alpha v_{t+1} \rho'_t}{M_F}\right)^{\frac{1}{1-\alpha}} \bar{Z}_t \tag{1.8}$$

[8] $\varepsilon^n_{f,t}$ and $\varepsilon^m_{f,t}$ are zero mean random variables bounded from below, such that $\frac{\Delta z^N_{f,t}}{z_{f,t}}$ and $\frac{\Delta z^M_{f,t}}{\bar{Z}_t}$ are always positive.

[9] This assumption keeps the number of R&D workers constant in general equilibrium while the number of goods can grow at a constant rate. Moreover, the endogenous growth rate of the economy is independent of the population size under this assumption.

$$v_t = \frac{P_t C_t}{\sigma} - M_F \left(\frac{A_N \alpha v_{t+1} \rho_t \prime}{M_F} \right)^{\frac{1}{1-\alpha}} + \rho' v_{t+1} \left[1 + A_N \left(\frac{A_N \alpha v_{t+1} \rho_t \prime}{M_F} \right)^{\frac{\alpha}{1-\alpha}} \right] \quad (1.9)$$

where

$$\rho_t' = \frac{\rho I_t C_t P_t}{I_{t+1} C_{t+1} P_{t+1}}.$$

(1.7) ((1.8)) equates the marginal cost of innovation (imitation) to the expected marginal return from innovation (imitation). A firm's optimal labor input in innovation $N_{f,t}$ is proportional to the firm's private knowledge $z_{f,t}$; and the labor input in imitation $M_{f,t}$ is proportional to public knowledge $\bar{Z}_t$. Equation (1.9) means that the marginal value of current market share is the current marginal profit plus the discounted future profit from innovation.

Notice that v in (1.9) is the private return of knowledge capital accumulation, which is smaller than the social return of knowledge, due to knowledge externality through imitation. The externality is captured by $\frac{I_t}{I_{t+1}}$ in ρ': a higher growth rate of number of goods due to larger imitation productivity A_M erodes an existing product's future market share.

The constant component of the firm value function u_t is given by

$$u_t = -\left(\frac{A_M \alpha v_{t+1} \rho_t'}{M_F} \right)^{\frac{1}{1-\alpha}} + \rho_t' \frac{I_t}{I_{t+1}} u_{t+1} + \rho_t' \frac{1}{\alpha} \left(\frac{A_M \alpha v_{t+1} \rho'}{M_F} \right)^{\frac{1}{1-\alpha}}. \quad (1.10)$$

u_t is equal to the expected discounted future profit from the imitated products. In other words, u_t measures the public knowledge pool's externality to each firm. In the equilibrium with free entry, u_t must be smaller than or equal to the fixed entry cost F. Since potential entrants are firms with zero products, their expected profit from entering is purely the public knowledge externality u_t. When the externality u_t is greater than entry cost F, new entrants will keep entering and diluting the public knowledge pool $\bar{Z}_t$. The average firm size $\bar{Z}_t$ will shrink until u_t is equal to F and no more entry occurs. Substituting $u_t = F$ into (1.10) shows that the total number of firms M_F decreases with entry cost F.

Firm f's period t dividend is

$$D_t = \frac{P_t C_t}{\sigma} \frac{z_{f,t}}{I_t} - \left(\frac{A_N \alpha v_{t+1} \rho_t'}{M_F} \right)^{\frac{1}{1-\alpha}} \frac{z_{f,t}}{\bar{Z}_t} + \left(\frac{A_M \alpha v_{t+1} \rho_t'}{M_F} \right)^{\frac{1}{1-\alpha}}.$$

Negative dividend means that firm f finances the R&D cost from shareholders at zero interest rate.

Firm Size Dynamics

Substituting (1.7) and (1.8) to (1.3), (1.4), and (1.5), the firm size dynamic process in (1.3) can be summarized by

$$z_{f,t+1} = R_{f,t+1} z_{f,t} + L_{f,t+1}, \tag{1.11}$$

where

$$R_{f,t+1} \equiv \frac{I_t}{I_{t+1}} \left(1 + A_N^{\frac{1}{1-\alpha}} \left[\frac{\alpha v \rho_t'}{M_F} \right]^{\frac{\alpha}{1-\alpha}} \right) + \varepsilon^n_{f,t+1}, \tag{1.12}$$

$$L_{f,t+1} \equiv \frac{I_t}{I_{t+1}} A_M^{\frac{1}{1-\alpha}} \left[\frac{\alpha v \rho_t \prime}{M_F} \right]^{\frac{\alpha}{1-\alpha}} + \varepsilon^m_{f,t+1}.$$

I decompose a firm's expected growth rate $g_{f,t}$ into an innovation rate $r_{f,t}$ and an imitation rate $l_{f,t}$.

$$E\left(r_{f,t}\right) \equiv \frac{E\left(\Delta z^N_{ft}\right)}{z_{f,t}} = A_N^{\frac{1}{1-\alpha}} \left[\frac{\alpha v \rho_t'}{M_F} \right]^{\frac{\alpha}{1-\alpha}} \tag{1.13}$$

$$E\left(l_{f,t}\right) \equiv \frac{E\left(\Delta z^M_{ft}\right)}{z_{f,t}} = A_M^{\frac{1}{1-\alpha}} \left[\frac{\alpha v \rho_t \prime}{M_F} \right]^{\frac{\alpha}{1-\alpha}} \frac{\bar{Z}_t}{z_{f,t}} \tag{1.14}$$

The expected innovation rate $E\left(r_{f,t}\right)$ is a constant and independent of firm size $z_{f,t}$; but the expected imitation rate $E\left(l_{f,t}\right)$ is scale-dependent. As firm f grows larger, its imitation rate declines simply because the public knowledge pool $\bar{Z}_t$ becomes smaller relative to firm f's size $z_{f,t}$.

In total, the expected firm growth rate $E\left(g_{f,t}\right) \equiv E\left(r_{f,t}\right) + E\left(l_{f,t}\right)$ declines with firm size purely because of the scale-dependent imitation rate.

$$E\left(g_{f,t}\right)=A_N^{\frac{1}{1-\alpha}}\left[\frac{\alpha v\rho_t'}{M_F}\right]^{\frac{\alpha}{1-\alpha}}+A_M^{\frac{1}{1-\alpha}}\left[\frac{\alpha v\rho_t\prime}{M_F}\right]^{\frac{\alpha}{1-\alpha}}\frac{\bar{Z}_t}{z_{f,t}}.$$

Moreover, the expected imitation rate $E\left(l_{f,t}\right)$ declines faster (or is more scale-dependent) when cross-firm knowledge spillovers are more efficient (A_M is greater) in (1.14). As a result, the firm's expected growth rate $E\left(g_{f,t}\right)$ also drops faster in a sector with more abundant knowledge spillovers than in other sectors.

Model Implication 1: A firm's imitation rate declines with firm size while a firm's innovation rate is independent of firm size. Moreover, a firm's imitation rate drops faster in sectors with more abundant knowledge spillovers than in other sectors, which causes the cross-sector difference in the scale-independency of firm growth rate.

Notice that (1.13) and (1.14) also provide insights into the sector-specific optimal research and development policy. There are two types of R&D: innovation that relies on intra-firm knowledge diffusion, and imitation that depends on inter-firm knowledge diffusion. Moreover, both R&D outputs have increasing return to their productivity A_M and A_N ($\frac{1}{1-\alpha}$ and $\frac{1}{1-\beta}$ >1). If knowledge diffuses faster within a firm than across firms ($A_N > A_M$) and $\alpha=\beta$, increasing A_N by 1% causes a greater growth rate increment than increasing A_M by the same amount and vice versa. The reason is that firms endogenously allocate more R&D input to the type with a comparative advantage in knowledge diffusion.

Take Natural Gas and Petroleum industry as an example. Since the oil land in Alberta Canada contains a different chemical ingredient from the oil land in Texas U.S., firms in this industry may find other firm's technology useless in their production or research. A strengthened intellectual property rights policy that encourages firms to use their private knowledge in R&D is more effective than policies that facilitate the sharing of information between firms. In Computer and Office Accounting Machinery industry, however, firms obey a common industrial standard and technologies are codified in most cases. A R&D policy that encourages firms to share information is more suited. In summary, a tailored sector-specific policy that favors the R&D type that allows for more efficient knowledge diffusion helps to achieve a higher economic growth rate.

Policies to support imitation (increase A_M) include subsidizing cross-firm R&D cooperation, facilitating labor turnover, encouraging universities to disseminate knowledge to the public, and so on. Strict intellectual property rights protection supports innovation (increase A_N).

1.2.3 Determinants of Firm Size Distribution

To provide economic context for (1.11), I want to compare it with an AR(1) process. $R_{f,t}$ here is a random variable while for a typical $AR(1)$ process R is a constant. In an $AR(1)$ process $z_{t+1} = Rz_t + L_t$, R measures persistency and L represents the randomness of the stochastic process. Similarly, in (1.11) $R_{f,t}$ measures the persistency of firm size, or to what extent current firm size affects future firm size by providing private knowledge capital for future innovation. $L_{f,t}$ indicates how much firms can learn from public knowledge capital which is independent of current firm size. If R is estimated using panel data of firm size controlling for firm fixed effects, it varies between 0.6 to 0.96 across sectors. Moreover, the estimated R is higher in a sector with more a heterogeneous firm size distribution, which confirms that a more persistent firm size dynamics is associated with a more dispersed size distribution.

Imagine an economy without imitation, which means eliminating $L_{f,t}$ in (1.11). Starting from a sector with many equally sized firms, and repeating the process for $z_{f,t+1} = R_{t+1}z_{f,t}$ for numerous periods, firms will end up with different sizes because they have different 'luck' in their innovation history. Overtime, firm size dispersion will grow without bound. The volatility of innovation shocks $\varepsilon^n_{f,t}$ determines how quickly the size dispersion explodes. In the real world with chances to learn from other firms, $L_{f,t}$ constrains and attenuates the size dispersion generated by innovation shocks. In equilibrium, firm size heterogeneity measured by the standard deviation (s.d.) of log-scale firm size is constant over time with imitation.

Overtime, the firm size distribution moves forward with the lower bound of the distribution rising at a constant growth rate. For a Pareto distribution, the shape parameter that determines the firm size dispersion does not change with the lower bound of the distribution. Therefore, even with a steady growth rate of the average firm size, the firm size distribution maintains its shape.

Proposition 1 *According to theorem 5 in Kesten (44), the firm size distribution* $\left\{z_{f,t}\right\}$ *in a given sector follows a Pareto distribution with scale parameter* μ*, such that* $E\left(R_{f,t}\right)^{\frac{1}{\mu}}=1$*, if* $\{R_{f,t},L_{f,t}\}$ *in the market size dynamics (1.11) are independently and identically distributed over time and across firms*[10]

.

Lemma 2 *When* $\{\log\left(R_{f,t}\right)\}$ *follows a normal distribution with variance* σ_r^2*, and* $\{R_{f,t},L_{f,t}\}$ *in the market size dynamics (1.11) are independently and identically distributed over time and across firms, there is a closed form solution for* μ*:*

$$\mu = 1-\frac{2ln\left\{E\left(R_{f,t+1}\right)\right\}}{\sigma_r^2}\approx 1+\frac{2\frac{\bar{l}}{1+\bar{r}+\bar{l}}}{\sigma_r^2},$$

$$\bar{l}\equiv\frac{1}{M_F}\sum_{f=1}^{M_F}r_{f,t},\ \bar{r}\equiv\frac{1}{M_F}\sum_{f=1}^{M_F}l_{f,t}.$$

$\bar{l}$ ($\bar{r}$) is the cross-firm average imitation (innovation) rate in the sector. The number of goods growth rate is defined as $g=\bar{l}+\bar{r}$. Over time, the average firm size $\bar{Z}_t$ keeps growing at a constant rate g, but the size dispersion measure $\frac{1}{\mu}$ is a constant.

(2) highlights two offsetting forces shaping firm size distribution: the innovation shock's volatility creates firm size difference while imitation reduces the difference. As mentioned in the last paragraph, the innovation shock's volatility σ_r^2 determines how quickly firm size dispersion explodes without imitation. On the other hand, imitation's relative contribution to the gross growth rate $\frac{\bar{l}}{1+\bar{r}+\bar{l}}$ defines the power of mean reversion to constrain the firm size dispersion from exploding. In total, firm size heterogeneity[11] $\frac{1}{\mu}$ declines with the relative magnitude between these two offsetting forces. Since abundant cross-firm knowledge spillovers or high A_M increase imitation's relative contribution to the gross growth rate $\frac{\bar{l}}{1+\bar{r}+\bar{l}}$,

[10] The independent assumption is unnecessary according to Goldie (29).

[11] For a *Pareto* distribution with scale parameter μ, $\frac{1}{\mu}$ is equal to the standard deviation of log scale firm size, which is commonly used as a measure of firm size heterogeneity in the literature.

a sector with more abundant cross-firm knowledge spillovers also exhibits more homogeneous firm size distribution than other sectors.

Model Implication 2: Firm size heterogeneity declines with the relative magnitude between imitation's gross growth rate contribution and innovation risk's volatility $\frac{\frac{\bar{l}}{1+\bar{r}+\bar{l}}}{\sigma_r^2}$.

Model Implication 3: Sectors with more abundant cross-firm knowledge spillovers have more homogeneous firm size distribution than other sectors.

1.2.4 Growth Rate Volatility Decomposition

$$Var(g_{f,t}) = Var(R_{f,t}) + \frac{Var(L_{f,t})}{(z_{f,t})^2} \tag{1.15}$$

According to the market share dynamics (1.11) and (1.12), every firm's growth rate is subject to two shocks: innovation risk and imitation risk. The relative weight of these two risks in a firm's growth volatility is different across firms. For a large firm, the main risk component is innovation risk, while for a small firm, the major component is imitation risk. Overall, a firm's growth volatility declines with its size, because innovation risk is the same across firms, while imitation risk's contribution to total volatility decreases with firm size.

Decomposing firm volatility into innovation risk and imitation risk sheds light on recent discoveries about the converging firm growth volatility among small private firms and large public firms. Comin and Mulani (16) and Davis, Haltiwanger, Jarmin, and Miranda (19) find that U.S. large public traded firms' volatility has risen, while small private firms' volatility has declined over the last several decades. One possible mechanism to explain these two concurrent facts is that certain policy or technology changes encouraged firms to invest more in innovation and less in imitation (A_N rises and/or A_M decreases). As a result of this policy change, firms undertake riskier projects, allocating more generous funds to innovation; the opposite happens when firms choose imitation projects requiring more limited funds. Such changes induce innovation volatility $Var(R_{f,t})$ to rise and imitation volatility $Var(L_{f,t})$ to drop at the same time. Since $Var(R_{f,t})$ is the major risk component for large firms and $\frac{Var(L_{f,t})}{(z_{f,t})^2}$ is the major risk component for a small firm, the increment of $Var(R_{f,t})$ dominates the decrement of $\frac{Var(L_{f,t})}{(z_{f,t})^2}$ for a large firm's volatility; meanwhile the decrement in $\frac{Var(L_{f,t})}{(z_{f,t})^2}$ outweighs the increment of $Var(R_{f,t})$ for small firms.

There is existing literature on declining knowledge spillovers, which deter imitation and encourage innovation. Caballero and Jaffe (9) and Rosell and Agrawal (62) find that the potency of spillover from old ideas to new knowledge generation has been declining over the last century. The policy changes started with the Bayh-Dole Act (35 USC 200-212) 1980, which grants patents to inventors who are funded by federal assistance. Since then, U.S. Patent Law has been amended several times to include increasingly broad infringement definitions. These policy changes all encourage innovation and limit imitation. Even universities, whose traditional role was to disseminate knowledge, have become more and more commercially oriented. Another related discovery is the divergence of moderating aggregate volatility and rising firm-level volatility for publicly traded firms. Comin and Mulani (16) propose an explanation also based on changing R&D activity: firms spend more resources on Embodied innovations and less on Disembodied innovations. The first type of R&D is patentable, so firms can appropriate all the benefits it generates. The second type of R&D is difficult to patent and easy to reverse engineer. The firm that develops a disembodied innovation cannot appropriate the benefits enjoyed by other firms when adopting it. The co-movement across firms weakens when there are fewer disembodied innovations to be imitated by everyone simultaneously. Since total output volatility is the summation of individual firms' volatility and the covariance between firms' growth rates, weaker co-movement reduces GDP volatility.

Summating the firm dynamics in (1.11) to the aggregate level may provide a coherent understanding of both the volatility convergence between large and small firms at firm level and the moderating aggregate volatility at the macro level. The key is firms' changing R&D patterns: less imitation and more innovation.

1.2.5 General Equilibrium

In general equilibrium, the average firm value, $\frac{v}{M_F}$, the growth rate in the number of goods, g, and the number of firms, M_F are solved using (1.16) to (B..12).

$$\left(1-\frac{\rho}{1+g}\right)\frac{v}{M_F}=\frac{PC}{M_F}+\frac{1-\alpha}{\alpha}\left(\frac{A_N\rho\alpha v}{(1+g)M_F}\right)^{\frac{1}{1-\alpha}} \tag{1.16}$$

$$g=\left(A_N^{\frac{1}{1-\alpha}}+A_M^{\frac{1}{1-\alpha}}\right)\left[\frac{\rho\alpha v}{(1+g)M_F}\right]^{\frac{\alpha}{1-\alpha}} \tag{1.17}$$

$$F=u=\frac{1-\alpha}{\alpha\left(1-\frac{\rho}{1+g}\right)}\left(\frac{A_M\rho\alpha v}{(1+g)M_F}\right)^{\frac{1}{1-\alpha}} \tag{1.18}$$

$$PC=L+\frac{PC}{\sigma}-\frac{\rho\alpha vg}{1+g} \tag{1.19}$$

In (B..12), the average firm value $\frac{v}{M_F}$ increases with entry cost F, but decreases with imitation productivity A_M, because v is the private return of knowledge capital, which shrinks with larger externality. In (1.16) to (1.17), higher A_M imposes two conflicting effects on growth rate g: first, it raises the imitation R&D input for the given average firm value $\frac{v}{M_F}$; second, it reduces average firm value $\frac{v}{M_F}$ for given R&D input because emerging new products squeeze the current products' market share. Higher A_N, however, always boosts economic growth, because it increases both innovation input and the private return of knowledge v.

(B..13) is the consumer's budget constraint. $\frac{PC}{\sigma}-\frac{\rho\alpha vg}{1+g}$ is the total dividend. The total R&D labor input $\frac{\rho\alpha vg}{1+g}$ increases with the private return of knowledge capital v, consumer's patience ρ, labor share in knowledge production function α and growth rate g.

Note that the economic growth rate, or the total number of goods growth rate, g, is independent of population size L. As L enlarges, market size PC and marginal firm value v increase proportionally. In the mean time, a larger market also accommodates more firms as indicated in (B..12), which means $\frac{v}{M_F}$ remains unchanged.

This model allows policy to affect economic growth. Suppose IPRP policy can not change innovation efficiency A_N, but it can constrain a firm's utilization of public knowledge and reduce A_M. Define $\gamma \equiv \frac{A_M}{A_N}$, which is smaller if IPRP is stronger. A better IPRP on one hand reduces imitation rate and harms growth rate; on the other hand, it raises the private value of knowledge capital v and indirectly increases economic growth. The growth maximizing γ is determined by the balance of these two factors.

1.3 Empirical Results

Before testing the three implications listed above, I introduce the data briefly.

1.3.1 Data

The NBER Patent Citation Data comprise detailed information on almost three million U.S. patents granted between January 1963 and December 1999, more than 16 million citations made to these patents between 1975 and 1999, and around 20,000 patent assignees, 92% of which are non-governmental organizations. I refer to all the organizations as 'firms' henceforth. Each patent contains highly detailed information on the innovation itself, the inventors, the assignee, etc. Moreover, patents have very wide industry and geographic coverage. The patents are classified to 42 wide SIC (Standard Industrial Classification) sectors. The percentage of U.S. patents awarded to foreign inventors has risen from about 20% in the early 1960s to about 45% in the late 1990s[12].

The citation data is well suited to this paper's purpose because these citations provide detailed paper trails of intellectual interactions across firms and sectors. Aggregated by industry level, the average values of time lag, the geographic distance, and the percentage of cross-firm citations indicate the pace and abundance of knowledge diffusion in each sector. Aggregated by firm level, cross-firm citations describe the sources of knowledge in each firm's R&D process. The firm-level aggregation allows for the distinction between imitation and innovation's contributions to each firm's overall growth rate, which is critical for testing the first two implications of the one-sector model. The industry aggregation allows for the testing of the model's third implication.

[12] See Hall, Jaffe, and Trajtenberg (31) for more details.

Here, I use patent citation to measure knowledge flow[13]. However, citations do not represent a one-to-one mapping of direct knowledge flows. A high proportion of noise may exist, because only some citations are made by the applicant, and others by the patent examiner. Jaffe, Trajtenberg, and Fogarty (39) and Duguet and MacGarvie (21) justify the use of aggregate patent citations as an indicator of knowledge spillovers based on a survey of patent inventors in the U.S. and firms in France. They conclude that some of the citations are associated with real knowledge flow, and patent citations aggregated at the industrial or regional level are valid measures of knowledge flow.

1.3.2 Implication 1: Scale-Independency of Firm Growth Rate

The one-sector model's first implication is: A firm's imitation rate declines with firm size, while a firm's innovation rate is independent of firm size. Moreover, a firm's imitation rate drops faster in a sector with more abundant knowledge spillovers than in other sectors, which causes the cross-sector difference in the scale-independency of firm growth rate.

In this subsection, I first demonstrate that the growth rate of a surviving firm has various scale dependencies in different sectors. I then attribute the above phenomenon to the scale-dependent imitation rate and its cross-sector differences.

Figure 1.3 and Figure 1.4[14] show that firm growth rate in the 'Petroleum and natural gas extraction and refining' sector is almost independent of firm size, while in 'Office computing and accounting machinery' it drops rapidly as firm size increases. Notice that the former sector has a more heterogeneous firm size distribution than the latter. In Figure 1.3 firm size (growth rate) is measured by French manufacturing firms' total revenues (growth rate of total revenues) in the Amadeus Database, while in Figure 1.4 firm size (growth rate) is measured by a firm's number of patents (growth rate of number of patents) in the NBER Patent Citation Data. In the model, the firm growth rate in terms of number of goods or total revenues is

[13]Patents cite other patents as 'prior art', with citations describing the property rights conferred. While a patent grants the assignee the right to exclude others from practising the invention described in the patent, it does not necessarily grant the owner the right to use the invention without the permission of cited assignees.

[14]The x-axis values are discounted by sector average so that the two sectors have similar domains in firm size.

the same.

For every sector, I run the following regression with both NBER Patent Citation Data and French manufacturing firm data:

$$g_{f,t} = a_{s,t} - b_{s,t} \ln\left(ps_{f,t}\right).$$

where $g_{f,t}$ is firm f's growth rate at time t. $ps_{f,t}$ is the number of patents granted to firm f by the beginning of time t (or firm f's total revenue at time t in the French firm data set). $a_{s,t}$ and $b_{s,t}$ are sector-specific. A larger $b_{s,t}$ means firm growth rate drops faster with firm size (or firm growth rate is more scale-dependent) in sector s at time t.

In Figure 1.5 (Figure 1.6), each scatter point represents one sector and the numbers label the four-digit NAICS (North American Industry Classification System) 2002 industry classification (SIC (Standard Industry Classification)). The firm size dispersion measure is the standard deviation of log scale firm revenue (patent stock) in Figure 1.5 (Figure 1.6). In both graphs, b_s declines with the firm size heterogeneity measure. In other words, firm growth rate is more scale-dependent in a sector with more homogeneous firm size distribution than other sectors.

This implication also predicts that when firm growth rate is broken down into innovation rate and imitation rate, the scale-independency of firm growth rate derives purely from the imitation rate (1.14).

A challenge in testing this implication is to estimate a firm's imitation rate $r_{f,t}$ and innovation rate $l_{f,t}$. Typically, we only observe a firm's overall growth rate, and it is difficult to tell what share is attributable to a firm's private knowledge and what share originates from public knowledge. The use of patent citations is a promising approach to solving this problem because they indicate the source of knowledge used during the patent invention. Within-firm citations (cross-firm citations) indicate that the citing firm uses its private (public) knowledge when creating a new patent.

At time t, firm f's patent stock growth rate $g_{f,t}$ is split into an innovation rate $r_{f,t}$ and imitation rate $l_{f,t}$, according to the ratio between within-firm citation and cross-firm citation[15].

$$\hat{g}_{f,t} = \frac{\text{No. of new patents}_{f,t}}{\text{patent stock}_{f,t}}$$

$$\hat{l}_{f,t} = \hat{g}_{f,t} \frac{\text{No. of cross-firm citations}_{f,t}}{\text{No. of total citations}_{f,t}}$$

$$\hat{r}_{f,t} = \hat{g}_{f,t} \frac{\text{No. of within-firm citations}_{f,t}}{\text{No. of total citations}_{f,t}}$$

Consider the following example. Firm f had ten patents at the beginning of year t. It obtained five new patents during year t. In these five patent applications, firm f's scientists cited 30 patents held by other firms and cited firm f's own patents 20 times. Firm f's patent stock growth rate at year t is $\hat{g}_{f,t} = \frac{5}{10} = 50\%$; the innovation rate is $\hat{r}_{f,t} = \hat{g}_{f,t} * \frac{30}{30+20} = 30\%$; and the imitation rate is $\hat{l}_{f,t} = \hat{g}_{f,t} * \frac{20}{30+20} = 20\%$.

In order to reflect the quality of the information transmitted in each citation count, I adjust the pure citation count by assigning a greater weight to a citation with shorter time lag or given by a more important citing firm. For example, if the citation time lag is n years, this citation is given a weight of $(1-\delta)^n$. δ is the knowledge capital depreciation rate. The three implications are virtually unaffected if I let the discount rate vary between 0 and 0.9. I use $\delta = 0.1$ in the following regressions. One reason to add time discount is that citations with shorter time lag transfer more frontier knowledge on average. The other reason is that firms usually cite inside patents sooner than outside patents. Without the time discount adjustment, I underestimate the inside knowledge flow and overestimate the outside knowledge flow. Since large firms cite themselves more intensively than small firms, without weighting citations by the importance of citing firm, I will underestimate the innovation rate for large firms and overestimate the imitation rate for small firms.

I run the following two regressions for every sector s and time t. Again, a larger $br_{s,t}$ ($bl_{s,t}$) means the innovation rate (imitation rate) is more scale-dependent in sector s at time t.

$$\hat{r}_{f,t} = ar_{s,t} - br_{s,t} \ln(ps_{f,t})$$

[15] I use only within-sector citations made and received by U.S. firms.

$$\hat{l}_{f,t} = al_{s,t} - bl_{s,t} \ln\left(ps_{f,t}\right)$$

Figure 1.7 shows the results for 1990. Similar patterns are exhibited in other years. Each scatter point represents one sector and the numbers label the SIC patent classification in the U.S. Patent Office. The imitation rate scale-dependencies $\left\{\hat{b}l_{s,t}\right\}$ for all sectors are around 0.1 to 0.3 and significantly different from 0 for all sectors. In contrast, innovation scale-dependencies $\left\{\hat{b}r_{s,t}\right\}$ are around 0 and 0.05[16] which are, for most sectors, not statistically significant. In addition, scale-dependency of imitation rate $\hat{b}l_{s,t}$ decreases with the sectoral firm size heterogeneity measure. This implies that the imitation rate is more scale-dependent in sectors with a more homogeneous firm size distribution, while scale-dependency of innovation rate $\hat{b}r_{s,t}$ is independent from the sectoral firm size heterogeneity measure.

In summary, when the growth rate is split into an innovation rate and imitation rate, the scale dependence of the growth rate comes only from the scale dependence of the imitation rate, since the innovation rate is independent of firm size. The firm growth rate drops faster in sectors with more homogenous firm size distribution because the imitation rate reduces more quickly in those sectors. Appendix 1.5 discusses why larger firms cite fewer outside patents and why firm growth rate declines with firm size.

1.3.3 Implication 2: Determinants of Firm Size Heterogeneity

The one-sector model's second implication is: Firm size heterogeneity declines with the relative magnitude between imitation's contribution to gross growth rate and innovation risk's volatility $\frac{\frac{\bar{l}}{1+\bar{r}+\bar{l}}}{\sigma_r^2}$.

For a *Pareto* distribution, the commonly used measure of firm size heterogeneity, i.e. standard deviation of log-scale firm size, is the reciprocal of the *Pareto* distribution scale parameter μ. (2) predicts that μ increases with imitation's contribution to gross growth rate, $\frac{\bar{l}}{1+\bar{r}+\bar{l}}$, and decreases with innovation shock's volatility, σ_r^2. In sector s and year t, $\bar{r}_{s,t}$ and $\bar{l}_{s,t}$ are given by the average of $\hat{r}_{f,t}$ and $\hat{l}_{f,t}$ for all

[16]The outlier, sector 51, has only 20 firms applying for patents in that year. $\hat{b}r$ is not statistically significant.

firms that applied for patents in sector s and year t; $\hat{\sigma}^2_{rs,t}$ is estimated by the standard deviation of $ln\left(\frac{1+\hat{r}_{f,t}}{1+\bar{r}_{s,t}+\bar{l}_{s,t}}\right)$ in (1.12); and sdlnps is the standard deviation of log-scale patent stock. Therefore, the model predicts that sdlnps increases in $\hat{\sigma}^2_{rs,t}$ and decreases in $\frac{\bar{l}}{1+\bar{r}+\bar{l}}$. $\frac{\bar{l}}{1+\bar{r}+\bar{l}}/\hat{\sigma}^2_{rs,t}$ is the relative magnitude of these two offsetting forces.

In Figure 1.8, each scatter point represents one sector and the numbers label the SIC patent classification in the U.S. Patent Office. The figure illustrates exactly what the model predicts. In this figure, the y-axis is the firm size heterogeneity measure s.d. of log scale patent stock and the x-axis is $\frac{\bar{l}}{1+\bar{r}+\bar{l}}/\hat{\sigma}^2_{rs,t}$. Therefore, the result supports the claim that when imitation's force dominates that of innovation risk's volatility, firm size distribution becomes more homogeneous.

1.3.4 Implication 3: Knowledge Spillovers Efficiency and Firm Size Heterogeneity

The one-sector model's third implication is: Knowledge spillovers are more abundant in sectors with more homogeneous firm size distribution.

I measure knowledge spillovers efficiency by the percentage of cross-firm citations among total citations and the citation time lag of cross-firm citations. The share of cross-firm citations among all citations indicates how likely it is that knowledge spillovers cross firm borders. Figure 1.9 shows that the proportion of cross-firm citations is negatively correlated with sectoral firm size heterogeneity.

The citation time lag, the interval between the application time of the citing patent and the application time of the cited patent, indicates the time needed for knowledge to travel between the inventors of these two patents. A shorter citation time lag for cross-firm citations indicates more efficient knowledge spillovers. Notice that the two inventors may take longer to exchange information if the geographic distance between them is larger. The great circle distance between the first inventor of the citing and the first inventor of the cited patent measures how far knowledge travels[17].

[17]The patent inventors are required to report their mailing address. From the Census 2000 U.S. Gazetteer Files, I identify over 90% of U.S. inventors' geographic locations by their five-digit ZIP code's latitude and longitude. Using both sides' latitude and longitude data, the great circle distance between the citing patent and the cited patent is calculated by the method in Sinnott (66).

Take the 'Office computing and accounting machinery' and 'Petroleum and natural gas extraction and refining' industries, for example. Figure 1.10 and Figure 1.11 show that knowledge diffusion is more likely to overcome firm borders faster in the former industry. Notice that the former sector has a more homogeneous firm size distribution than the latter sector. The gap between these two sectors shrinks as the time lag becomes longer, but still exists even after a lag of 20 years.

The fixed-effects OLS regressions in Table 1.1 give the determinants of cross-firm citation time lag with U.S. citations. Since the time lag of repetitive citations overestimates the knowledge spillovers time lag, I only include a citation the first time the citing firm cites the cited patent. First-time citations account for around 70% of all citations. The regression results are similar and more significant when all citations are included.

In the first regression, there are state pair fixed effects to capture time invariant unobserved variables that may have an impact on information diffusion between the citing state and the cited state. In the second regression, the sector fixed effects are included to take care of sector-specific time invariant elements that may affect within-sector knowledge spillovers. If sectoral firm size heterogeneity (s.d. of log(patent stock)) changes over time, the model implies that citation time lag should move in the same direction. In the third regression, both types of fixed effects are considered. In the fourth column, I control the citing firm and cited firm firm-pair fixed effects. Year dummies for the citing patent application year are included in all regressions.

In all regressions, the citation time lag is longer if geographic distance is larger, the citing organization is smaller, the cited organization is smaller, the sector size is smaller, or the sectoral firm size distribution is more heterogeneous. Distance delays the exchange of knowledge because it increases communication cost. Larger firms are quicker to acquire information, because they are on average older and have better connections due to a more established social network. A larger industry tends to have faster knowledge diffusion. Table 1.1 shows that the sectoral firm size heterogeneity has the predicted positive effect on citation time lag. One standard deviation change in s.d. of log(patent stock) (0.53) causes the citation time lag to increase by 0.44 (0.766*0.53) to 1.55 (2.93*0.53) years, keeping other conditions constant.

In summary, there is a greater proportion of cross-firm citations when the sectoral firm size distribution is more homogeneous. Additionally, among the cross-firm citations, citation time lag is shorter in sectors with more homogeneous firm size distribution, controlling for the size of the citing and cited firms, the size of the sector, and state-pair, sector and firm-pair fixed effects. These results support the third implication of the theoretical model: knowledge spillovers are more abundant in sectors with a homogeneous firm size distribution.

1.4 Multi-Sector Model

1.4.1 Facts about Multi-Sector Firms

When firm size is measured by the number of patents, firm size distribution within each sector follows a distinct *Pareto* distribution with scale parameter ranging from 0.29 to 3[18]. When all the patenting firms are pooled, the firm size distribution also follows a *Pareto* distribution with a scale parameter close to 1.68. Note that one firm may apply for patents in multiple sectors; the firm size in the pooled distribution of the whole economy is the summation of its number of patents in all sectors. This result corroborates the stylized facts in Helpman, Melitz, and Yeaple (33), with firm size measured by number of employees. In their paper, every sector s follows a *Pareto* firm size distribution with special scale parameter μ_s, while the aggregate economy also follows a *Pareto* firm size distribution with scale parameter μ close to 1. To this point, no research has been conducted to explain the universal *Pareto* distribution of firm size in each sector and in the entire economy.

The following phenomena inspire me to consider firm size dynamics from a multi-sector perspective. First, Table 1.4 and Table 1.5 show that many firms develop products in multiple sectors. Moreover, larger firms operate in more sectors. Table 1.4 is borrowed from Broda and Weinstein (8)[19], which highlights the multi-product nature of firms in these markets. It demonstrates that firms with higher

[18] Estimated by French Manufacturing Firm Data from Bureau van DIJK's Amadeus Database

[19] Streitweiser (67), Jovanovic (43) and Bernard, Redding, and Schott (7) also found a similar extent of industry diversification in U.S. firms or plants. UPCs in the second column means Universal Product Codes, commonly referred to as bar codes. Share in the last column means the total market share of the firms within each group.

sales in dollar value also sell a greater number of goods and sell in more sectors. Table 1.5 shows a similar result in patent data: organizations that own more patents also apply for patents in more patent categories.

Second, in the NBER Patent Citation Data, 37% of all citations are cross-sector Citations; the percentage becomes higher and approaches 100% when the sector division is finer. This suggests that knowledge spillovers exist not only within but also across sectors. In Table 1.6, the row index represents the citing sector, and the column index represents the cited sector. The (i, j) element of the matrix is the percentage of citations given by sector j to sector i. There are 42 sectors in total, from which I selected 6 for illustration. Every sector gives a large proportion of citations to the patents in the same sector, but also allocates a small proportion of citations to patents in every other sector. In Table 1.7, I adjust the original percentage of cross-sector citation by the cited sector's patent shock share in the data set of a given year. The table shows that every sector cites itself over-proportionally and cites other sectors under-proportionally in most cases, but there are some sectors that receive over-proportional citations (the blue cells). These blue cells indicate that the cited sector contributes above-average knowledge to the citing sectors.

Third, inter-sector knowledge spillovers cause firm size dynamics in each sectors to be dependent on each other and generate a pooling firm size distribution similar to the data. The multi-sector model expands upon the one-sector model by adding inter-sector knowledge spillovers. With inter-sector knowledge spillovers, a firm's growth dynamics in every sector and in the entire economy are subject to the impacts of all sectors. The similarity in growth dynamics confirms that firm size distributions, whether measured within one sector or in the entire economy, all converge to the *Pareto* distribution. The one-sector model is a special case of the multi-sector model, when cross-sector knowledge spillovers do not exist and sectors are independent.

Without inter-sector knowledge spillovers, firm growth dynamics in different sectors would be independent. The summation of several independent *Pareto* distributed variables is still *Pareto* distributed, but the scale parameter is equal to the minimum of the component distribution scale parameters[20]. In contrast, the firm-

[20] See Jessen and Mikosch (40) and Gabaix (27).

or establishment-level data (Figure 1.1 and Figure 1.2) show that the scale parameter of all firms' distribution lays between the component sectors' scale parameter values. Therefore, the firm size dynamics in different sectors must be dependent.

1.4.2 Model

A representative firm f operates in K sectors. Firm f's size in all sectors at time t is summarized by a K-dimensional real vector $z_{f,t}$. The k^{th} element of $z_{f,t}$, $z^k_{f,t}$, represents the number of products in the k^{th} sector invented by firm f. Firm f can apply its private knowledge capital in sector i, $z^i_{f,t}$, to the innovation in any sector j, where $i, j \in \{1, 2, ..., K\}$, using production function $A^{ij}_N \left(N^{ij}_{f,t}\right)^{\alpha} \left(z^i_{f,t}\right)^{1-\alpha}$. A^{ij}_N is the ability to apply sector i's knowledge to innovate in sector j (call it ij type innovation). $N^{ij}_{f,t}$ is firm f's research hours spent in ij's type of innovation. Firm f utilizes public knowledge capital, $\bar{Z}^i_t$, for imitation in any sector j with production function $A^{ij}_M \left(M^j_{f,t}\right)^{\beta} \left(\bar{Z}^i_t\right)^{1-\beta}$. A^{ij}_M is the ability to apply sector i public knowledge for imitation in sector j. $M^{ij}_{f,t}$ is firm f's research hours spent in ij's type of imitation.

Notice that the cross-sector knowledge spillovers happen both within firm borders and across firm borders. The $\left\{A^{ij}_N\right\}, i \neq j$ measures the cross-sector but within-firm-border knowledge spillovers efficiency, while $\left\{A^{ij}_M\right\}$ includes both the cross-sector and cross-firm knowledge spillovers. This assumption accords with the data for cross-firm citations, where cross-firm citations occur both within and across sectors. For simplicity, I assume $A_M = \gamma A_N$, where a smaller γ represents a stronger protection of intellectual property rights.

One striking feature of the knowledge diffusion matrix A_N is its asymmetry. Some sectors contribute intensive knowledge spillovers to a large number of other sectors, for example, *Electronic components and accessories and communications equipment*, *Office computing and accounting machines* and *Professional and scientific instruments*; while other sectors are almost isolated from the rest of the economy, for example, *Ship and boat building and repairing* and *Railroad equipment*.

Given the heterogeneous sectoral knowledge contribution to the economy, any growth policy should consider its impact on the relative private returns among sectors, because firms allocate research efforts across sectors according to the relative private returns. More importantly, in the general equilibrium section of Appendix A.2, I show that the cross-sector resource allocation is important for growth, and that when resources are allocated according to a sector's knowledge contribution to the economy, the economy obtains a higher growth rate.

Firm f's manager chooses $2K^2$ types of R&D inputs $\left\{N_{f,t}^{ij},\, M_{f,t}^{ij}\right\}$ because $i, j \in \{1, 2, ..., K\}$. That way, expected marginal returns from the $2K^2$ types of R&D are equal to their marginal costs. Solving a similar but more complicated firm's problem than (1.2) (see Appendix A.2 for details), the dynamics of firm size in all K sectors can be summarized by

$$z_{f,t+1} = R_{f,t} z_{f,t} + L_{f,t}. \tag{1.20}$$

$R_{f,t}$ is a $K \times K$ random matrix, and the (i, j) element $R_{f,t}^{ij}$ measures the success rate when firm f uses its private knowledge from sector j to innovate in the creation of new products in sector i. $L_{f,t}$ is a K-dimensional random vector. The k^{th} element of $L_{f,t}$ is the number of imitated products in sector k which firm f has invented. For instance, the size dynamics of firm f's branch in sector k are

$$z_{ft+1}^{k} = R_t^{k1} z_{ft}^{1} + ... + R_t^{kk} z_{ft}^{k} + ... + R_t^{kK} z_{ft}^{K} + L_t^{k}. \tag{1.21}$$

$$L_t^{k} = L_t^{k1} \bar{Z}_t^{1} + ... + L_t^{kk} \bar{Z}_t^{k} + ... + L_t^{kK} \bar{Z}_t^{K}$$

Proposition 3 *If $\{z\}$ follows the dynamics in (1.20) and the random matrices $R_{f,t}$ and $L_{f,t}$ in (1.20) satisfy the restrictions in Kesten (44) (4.9), for any vector $x \in R^K$ and $|x| = 1$, there exist some μ such that $\{x'z\}$ follows a Pareto distribution with parameter μ.*

The above proposition means that a *Pareto* distribution exists in any subset of the economy. For example, when studying the firm size distribution of the k^{th} sector, pick $x = (0,...0,1,0,....0)$ with the k^{th} element equal to one and all others set to zero. When studying the size distribution of all firms in the entire economy, pick $x = \frac{1}{\sqrt{K}}(1,...1,1,1,....1)$, and here the total firm size is the summation of its branches size in all K sectors.

Besides the implication about universal *Pareto* distribution, the knowledge diffusion matrix A also determines the allocation of research resources across-sectors. The knowledge of a sector k is more valuable and sector k attracts more research investment, if sector k contributes more intensive knowledge spillovers to the entire economy (see Appendix A.2 for details).

The multi-sector model suggests one more reason to protect intellectual property rights. When policy makers strengthen intellectual property rights so that γ is smaller, the private return of knowledge accumulation in each sector becomes higher, but the return of major knowledge-contributing sectors increases more than other sectors. As a result, firms invest a greater share of their research fund in the major knowledge givers than in other sectors. Economic growth indirectly benefits from the better cross-sector R&D resource allocation. In addition to the trade-off between higher imitation rate and lower private return of knowledge in the one-sector model, the growth maximizing γ here also takes into account its impact on cross-sector resource allocation.

1.5 Conclusion

This paper employs knowledge spillovers to examine two questions about firm size distribution: Why is firm size heterogeneity different across sectors? and Why do firm size distributions follow dependent *Pareto* distributions in every subset of the economy?

The one-sector model answers the first question using sector-specific inter-firm knowledge spillovers efficiency. In sectors with abundant knowledge spillovers, firms invest more in imitation and less in innovation; therefore imitation contributes more substantially to the overall growth rate. Since every firm has an equal chance to learn from public knowledge, imitation has a stronger influence on smaller firms' growth rates, which leads to a declining firm growth rate with firm size. Faster catch-up of smaller firms generates a more homogeneous firm size distribution.

The one-sector model implies that knowledge spillovers are more abundant, firm growth rate declines faster with firm size, and imitation contributes more to the gross growth rate in sectors with more homogeneous firm size distribution. The model has three testable implications that are supported by NBER Patent Citation

Data. The advantage of this data set is that it keeps track of inter-firm knowledge spillovers, which allows for the measurement of the speed of knowledge diffusion and the separation of the share of the innovation and imitation rates in the overall growth rate of the firm.

To answer the second question, the multi-sector model improves upon the one-sector model with two additional features: firms develop products in multiple sectors, and cross-sector knowledge spillovers allow for dynamics to interact across all sectors. As a result, the firm growth dynamic in any subset of the economy evolves in a pattern similar to that of the whole economy. This induces a *Pareto* firm size distribution with different scale parameters in any subset of the economy.

At the aggregate level, the micro-founded models lead to policy suggestions relevant to intellectual property rights and trade. The one-sector model suggests that the optimal level of intellectual property rights protection depends on the trade-off between high imitation rate and lower private return of innovation. The multi-sector model suggests one more channel by which policies affect economic growth: the cross-sector R&D resource allocation. Strong intellectual property rights promote economic growth by encouraging firms to invest a larger share of their research fund in the intensive knowledge-contributing sectors, so that resource allocation is justified by each sector's knowledge externality to the economy. In a similar manner, Cai and Li (11) shows that trade cost harms economic growth by distorting the relative private returns of innovation between sectors of heterogeneous knowledge contributions to the economy and the cross-sector research resource allocation.

Table 1.1: OLS Regressions with U.S. Citations

Dependent variable: citation lag				
Independent variable	1	2	3	4
S. D. of log(ps)[a]	.776**	2.934**	1.262**	2.443***
	(.372)	(1.236)	(.563)	(.455)
Log(dist) [b]	.137***	.122***	.142***	.030***
	(.017)	(.013)	(.008)	(.010)
Log(PS_{citing})[c]	-.147***	-.110***	-.099***	.555***
	(.008)	(.021)	(.004)	(.063)
Log(PS_{cited})[d]	-.152***	-.080***	-.113***	-2.621***
	(.008)	(.014)	(.005)	(.146)
Log($PS_{industry}$)[e]	-.355***	-3.470***	-3.215***	1.208***
	(.014)	(.429)	(.125)	(.128)
Fixed effects	State pair	Sector	State pair by sector	Firm pair
No. of observations	1132505	1132505	1132505	1132505
No. of groups	2626	42	47375	719298
R square	.095	.093	.173	.855

[a]Standard deviation of log scale patent stock for all firms in the sector.
[b]Log scale great circle distance between the citing patent and the cited patent
[c]Log scale patent stock of the citing firm.
[d]Log scale patent stock of the cited firm.
[e]Log scale patent stock of the sector.

Table 1.2: Summary of Variables

Summary of Variables					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Log(dist)	1132640	6.711	1.784	0	9.460
Log(PS_{citing})	1132640	2.539	2.574	0	10.490
Log(PS_{cited})	1132640	2.769	2.618	0	10.280
Log($PS_{industry}$)	1132640	10.611	1.052	4.111	12.382
S. D. of Log(PS)	1132640	1.667	0.217	0.786	8.495

Table 1.3: Correlation between Variables

Variable	Lag	Log(dist)	Log(PS_{citing})	Log(PS_{cited})	Log($PS_{industry}$)
Lag	1				
Log(dist)	0.0304	1			
Log(PS_{citing})	-0.1211	0.0133	1		
Log(PS_{cited})	-0.1106	-0.0137	0.371	1	
Log($PS_{industry}$)	-0.0605	0.055	0.2123	0.2123	1
S. D. of Log(PS)	0.0511	-0.021	0.1906	0.2139	0.0942

Table 1.4: Table 2 in Broda and Weinstein (8)

TABLE 2
Average Firm Characteristic by Firm Size (2003Q4)

Value Bins (in $)	UPCs	Brands	Modules	Product Groups	Share
1- 100,000	3	1	1	1	0.00
100,000 - 1,000,000	10	2	3	2	0.02
1,000,000 - 5,000,000	33	5	6	3	0.05
5,000,000 - 10,000,000	69	7	10	4	0.04
10,000,000 - 50,000,000	139	11	20	7	0.17
50,000,000 - 100,000,000	386	22	50	14	0.10
100,000,000 - 500,000,000	713	37	71	18	0.33
500,000,000 - 1,000,000,000	1191	68	110	28	0.12
> 1,000,000,000	3431	182	246	54	0.16

Note: The share is based on total value of UPCs for firms within a bin compared to total value in 2003Q4
Ignore entry and exit of firms, only consider firms that exists in both periods t and t-1

Table 1.5: Multi-Sector U.S. Patent Owners

Number of Patents	Average Number of Patent Categories
1 - 10	1.34
11 - 100	3.89
101 - 1000	8.93
1001 - 10000	15.17
10000-	25.57
Source: NBER Patent Citation Data 1999	

Table 1.6: Share of Cross-Sector Citations: Example

	The Cited Sector						
The Citing Sector	%	1	2	6	7	8	9
	1	80.34	0.00	0.00	6.74	0.00	8.15
	2	0.33	38.59	0.33	0.66	8.70	0.00
	6	0.11	0.42	60.30	10.72	1.27	0.42
	7	0.46	0.41	5.16	58.46	4.52	14.06
	8	0.00	1.44	1.32	7.38	66.33	0.06
	9	1.09	0.30	0.30	14.68	0.24	67.73

Table 1.7: Intensity of Cross-Sector Citations: Example

	The Cited Sector						
The Citing Sector	%/%	1	2	6	7	8	9
	1	80.34	0.00	0.00	6.74	0.00	8.15
	2	0.33	38.59	0.33	0.66	8.70	0.00
	6	0.11	0.42	60.30	10.72	1.27	0.42
	7	0.46	0.41	5.16	58.46	4.52	14.06
	8	0.00	1.44	1.32	7.38	66.33	0.06
	9	1.09	0.30	0.30	14.68	0.24	67.73

Table 1.8: Example Sector Names

1	Food and kindred products
2	Textile mill products
6	Industrial inorganic chemistry
7	Industrial organic chemistry
8	Plastic materials and synthetic resins
9	Agricultural chemicals

Table 1.9: Correlation between Different Measures of Firm Size Heterogeneity - French 4-Digit NAICS Manufacturing Sectors 1997

Measure	sd(lnl)	sd(lny)	sd(lns)	sd(lnva)
sd(lnl)	1			
sd(lny)	0.964	0.961		
sd(lns)	0.961	0.998	0.1	
sd(lnva)	0.970	0.963	0.964	1

Table 1.10: Correlation between Different Measures of Firm Size Heterogeneity - Chilean 4-Digit SITC Manufacturing Sectors 1996

Measure	sd(lnl)	sd(lny)	sd(lns)	sd(lnva)
sd(lnl)	1			
sd(lny)	0.825	1		
sd(lns)	0.769	0.933	1	
sd(lnva)	0.715	0.910	0.896	1

Table 1.11: OLS Regressions - Random Citations

Dependent variable: citation lag[a]			
Independent variable	1	2	3
S. D. of Log(PS)	.424***	.160**	.168***
	(.059)	(.078)	(.066)
Log(dist)	.065	-.051	0.039
	(.042)	(.026)	(0.026)
Log(PS_{citing})	-.040***	-.002	-.001
	(.005)	(.002)	(.002)
Log(PS_{cited})	-.170***	-.108	-.160***
	(.010)	(.036)	(.007)
Log($PS_{industry}$)	-.175***	-3.603***	-3.341***
	(.014)	(.179)	(.096)
Year fixed effect	Yes	Yes	Yes
State pair fixed effects	Yes	No	No
Industry fixed effects	No	Yes	No
State pair - industry fixed effects	No	No	Yes
No. of observations	2120904	2120904	2120904
No. of groups	2626	42	47375

[a]Robust standard errors clustered by sector are reported in the brackets. Year dummies are included.

Table 1.12: Summary of Variables - Random Citations

Summary of Variables					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Citation lag	2120904	8.332	6.385	0	94
Log(dist)	2120904	7.043	1.1897	0	9.530
Log(PS_{citing})	2120904	3.473	2.695	0	9.649
Log(PS_{cited})	2120904	3.134	2.667	0	9.563
Log($PS_{industry}$)	2120904	10.645	1.055	4.111	12.382
S. D. of Log(PS)	2120904	1.700	.243	.787	8.495

Table 1.13: Correlation between Variables - Random Citations

Variable	Lag	Log(dist)	Log(PS_{citing})	Log(PS_{cited})	Log($PS_{industry}$)
Lag	1				
Log(dist)	-.009	1			
Log(PS_{citing})	-.0317	-.047	1		
Log(PS_{cited})	-.062	-.040	.172	1	
Log($PS_{industry}$)	.020	.078	.211	.219	1
S. D. of Log(PS)	.054	-.042	.190	.199	.048

Table 1.14: OLS Regressions - G7 Citations

Dependent variable: citation lag[a]				
Independent variable	1	2	3	4
S. D. of Log(PS)	.576	.492	1.99**	1.917***
	(.353)	(1.164)	(.774)	(.440)
Log(dist)	.078***	.122***	.104***	.044***
	(.020)	(.050)	(.022)	(.007)
Log(PS_{citing})	-.149***	-.134***	-.112***	.562***
	(.014)	(.008)	(.008)	(.050)
Log(PS_{cited})	-.183***	-.179***	-.149***	-2.464***
	(.044)	(.059)	(.051)	(.164)
Log($PS_{industry}$)	-.335***	-2.787***	-2.939***	1.215***
	(.089)	(.365)	(.373)	(.104)
Fixed effects	State pair	Sector	State pair-sector	Firm pair
No. of observations	2158761	2158761	2158761	2158761
No. of groups	49	42	1884	1238745
R square	.089	.098	.111	.805

[a]Robust standard errors clustered by sector are reported in the brackets. Year dummies are included.

Table 1.15: Summary of Variables - G7 Citations

Summary of Variables					
Variable	Obs.	Mean	Std. Dev.	Min.	Max.
Citation lag	2158761	6.866	5.010	0	95
Log(dist)	2158761	7.310	2.036	0	9.760
Log(PS$_{c}iting$)	2158761	3.321	2.778	0	10.491
Log(PS$_{c}ited$)	2158761	3.451	2.689014	0	10.280
Log(PS$_{i}ndustry$)	2158761	10.651	1.033	4.111	12.382
S. D. of Log(PS)	2158761	1.687	.231	.786	8.494

Table 1.16: Correlation between Variables - G7 Citations

Variable	Lag	Log(dist)	Log(PS$_{c}iting$)	Log(PS$_{c}ited$)	Log(PS$_{i}ndustry$)
Lag	1				
Log(dist)	0.038	1			
Log(PS$_{c}iting$)	-0.155	0.035	1		
Log(PS$_{c}ited$)	-0.162	0.028	0.423	1	
Log(PS$_{i}ndustry$)	-0.058	0.025	0.273	0.278	1
S. D. of Log(PS)	0.026	0.024	0.157	0.181	0.069

Figure 1.1: Firm Size Distribution in Different Sectors - French Manufacturing Firms

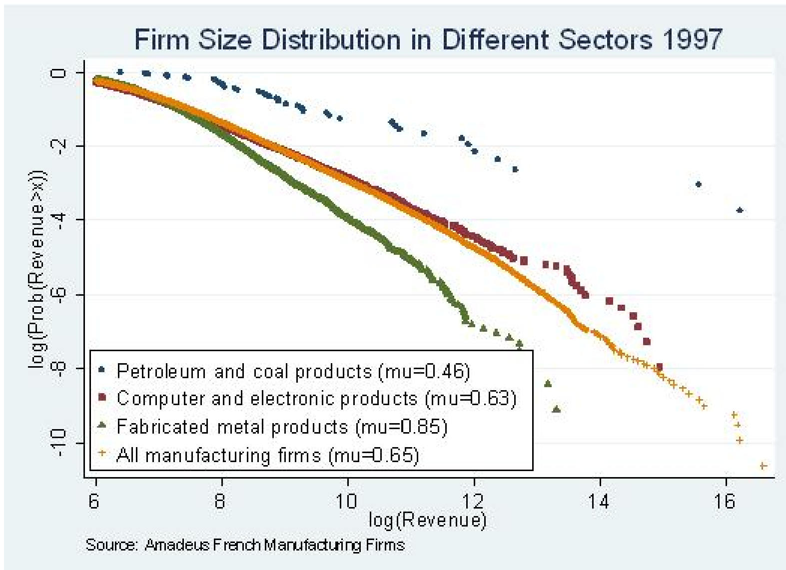


Figure 1.2: Firm Size Distribution in Different Sectors - U.S. Patent Owners

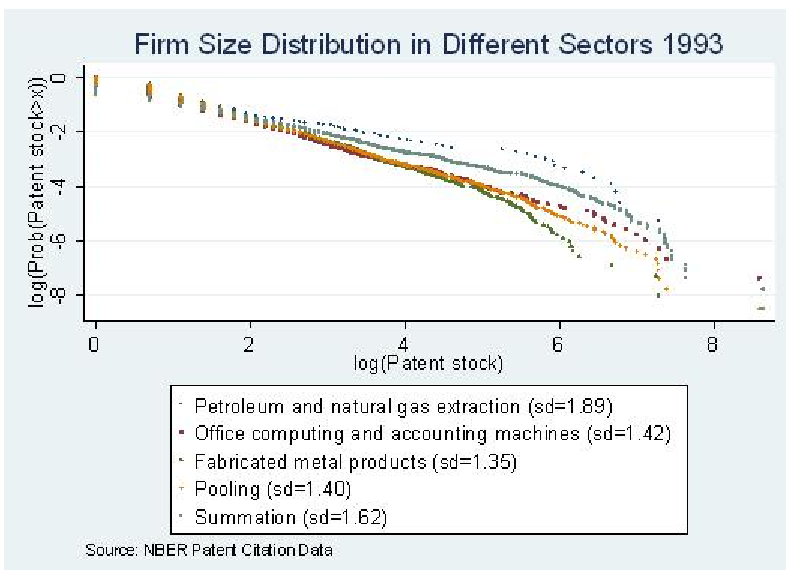


Figure 1.3: Five-Year Firm Growth Rate - French Manufacturing Firms

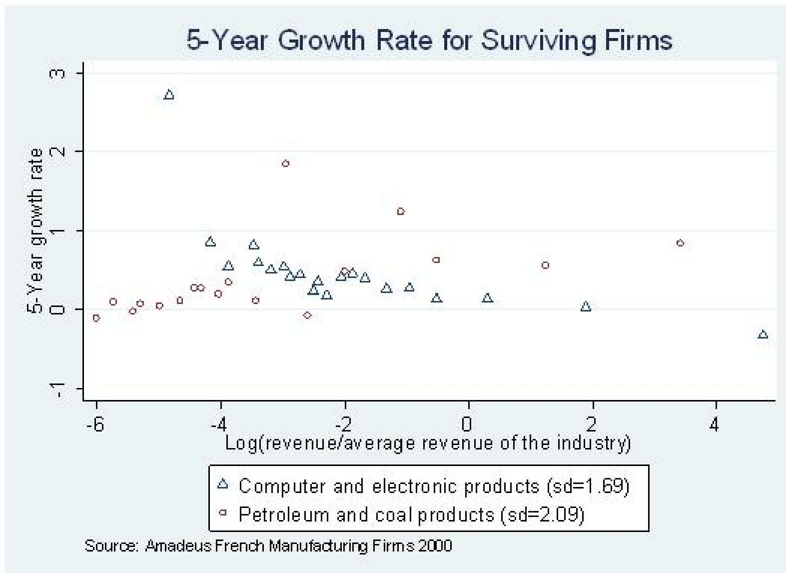


Figure 1.4: Five-Year Firm Growth Rate - U.S. Patent Owners

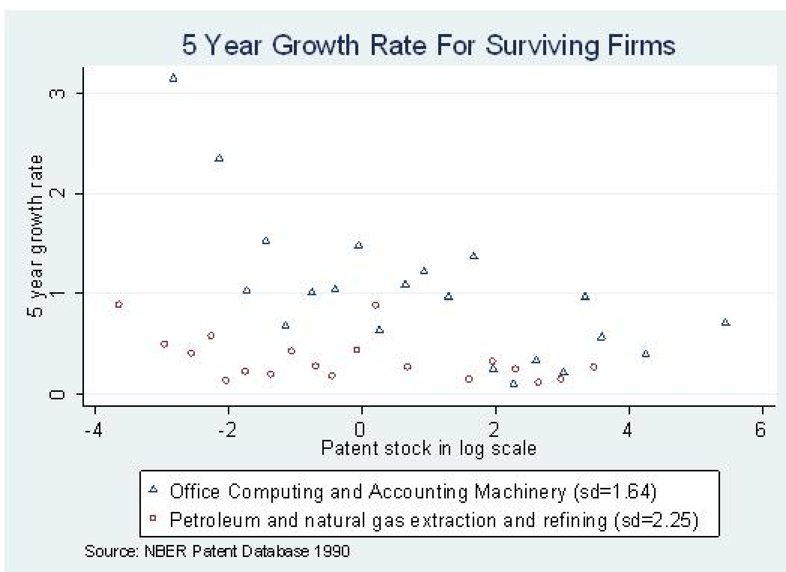


Figure 1.5: Scale-Dependency of Firm Growth Rate - French Manufacturing Firms

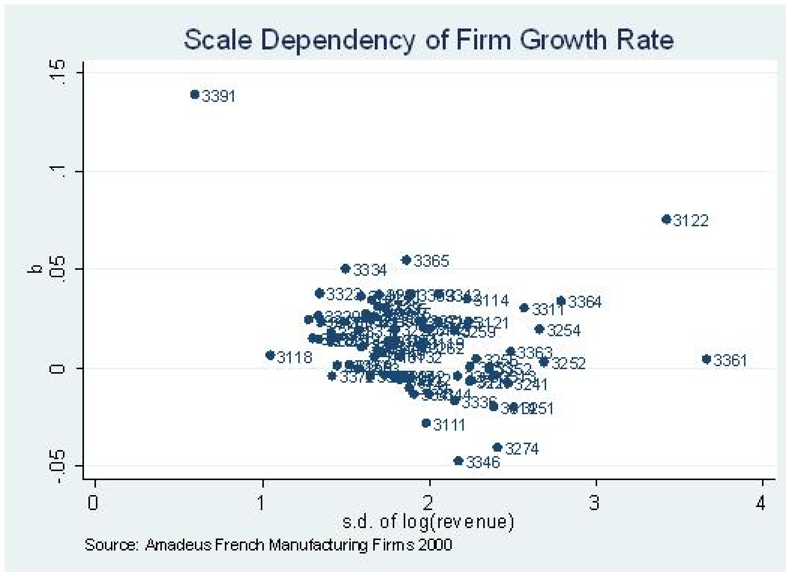


Figure 1.6: Scale-Dependency of Firm Growth Rate - U.S. Patent Owners

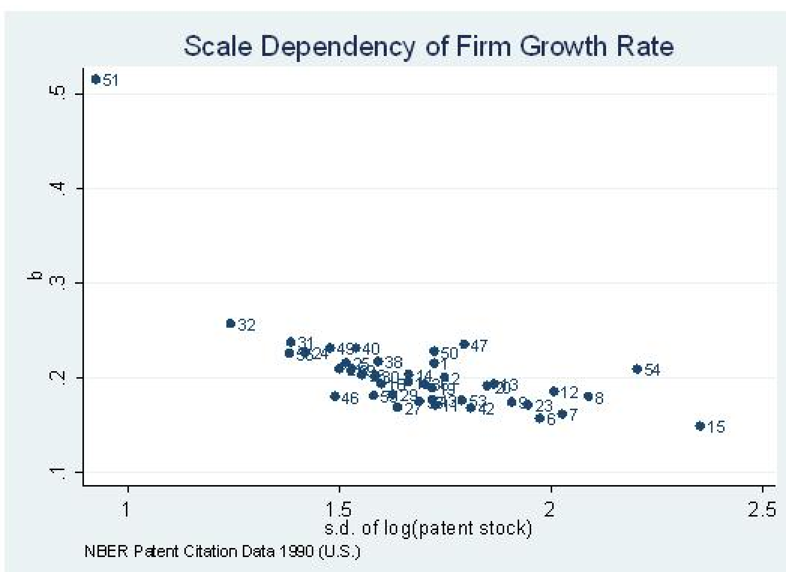


Figure 1.7: Scale-Dependency of Imitation Rate and Innovation Rate - French Manufacturing Firms

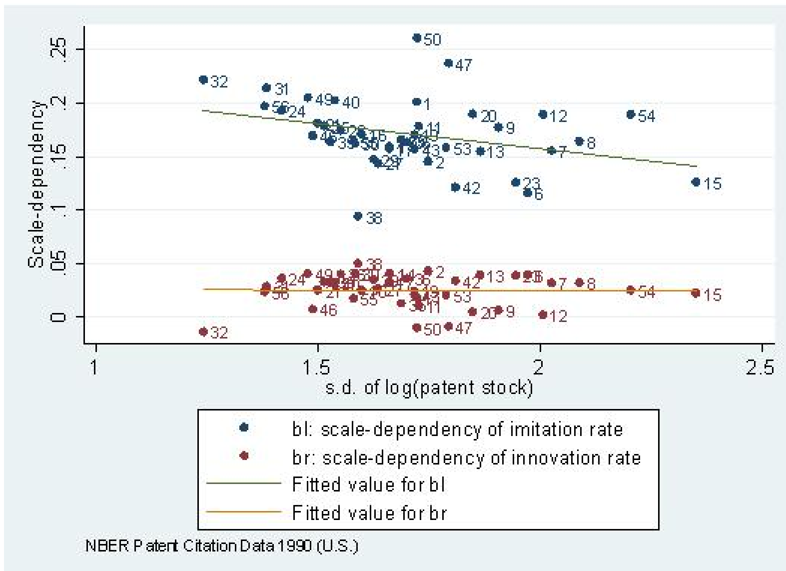


Figure 1.8: Scale-Dependency of Imitation Rate and Innovation Rate - U.S. Patent Owners

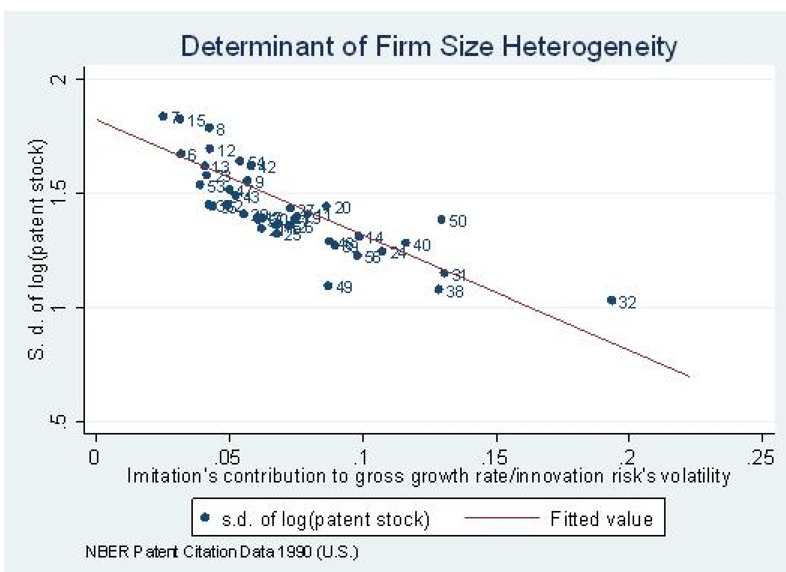


Figure 1.9: Cross-Firm Citation's Share in Total Citation by Sectors

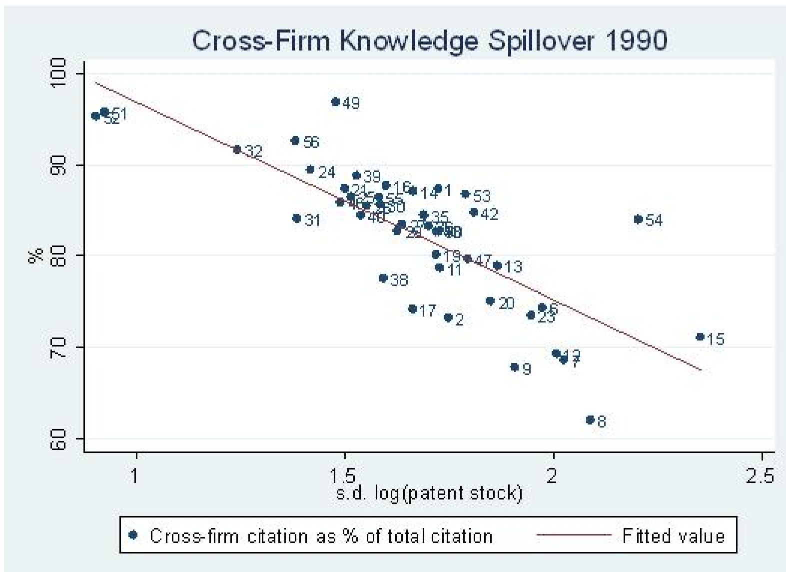


Figure 1.10: Cross-Firm Citation's Share in Total Citation by Year

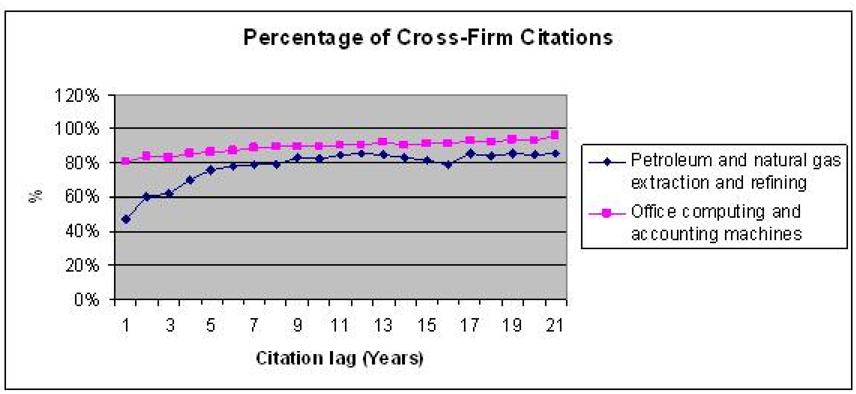


Figure 1.11: Average Citation Distance by Year

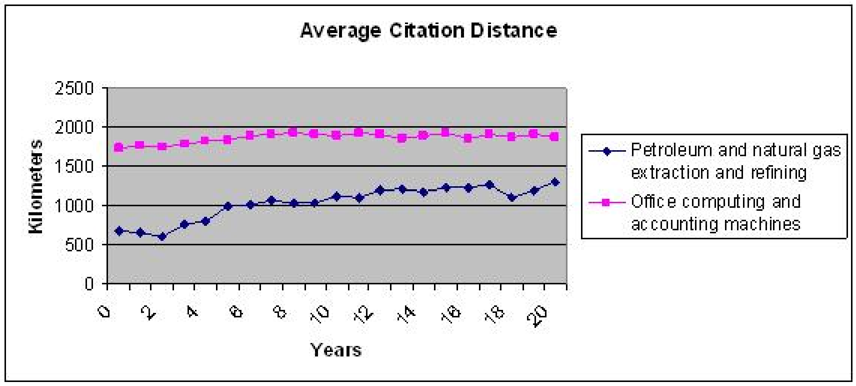


Figure 1.12: Sectoral Firm Size Heterogeneity in U.S. 1997 and 2002

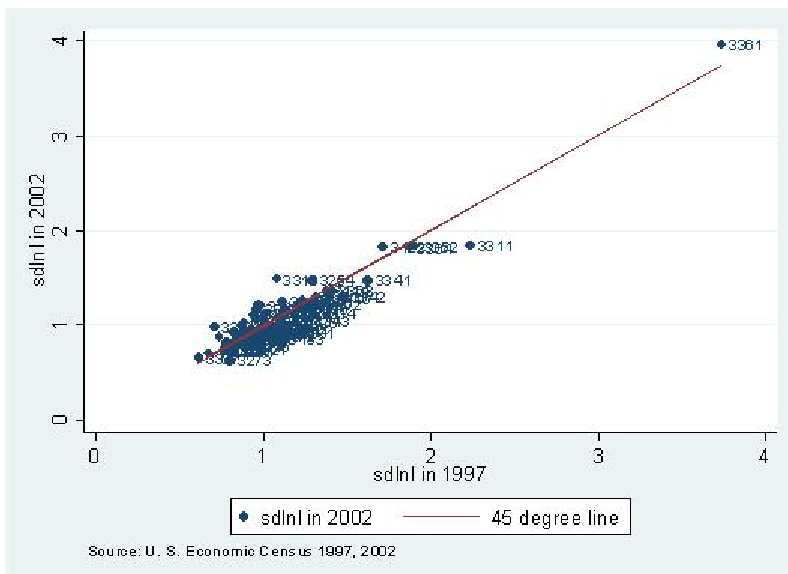


Figure 1.13: Sectoral Firm Size Heterogeneity in France 1997 and 2005

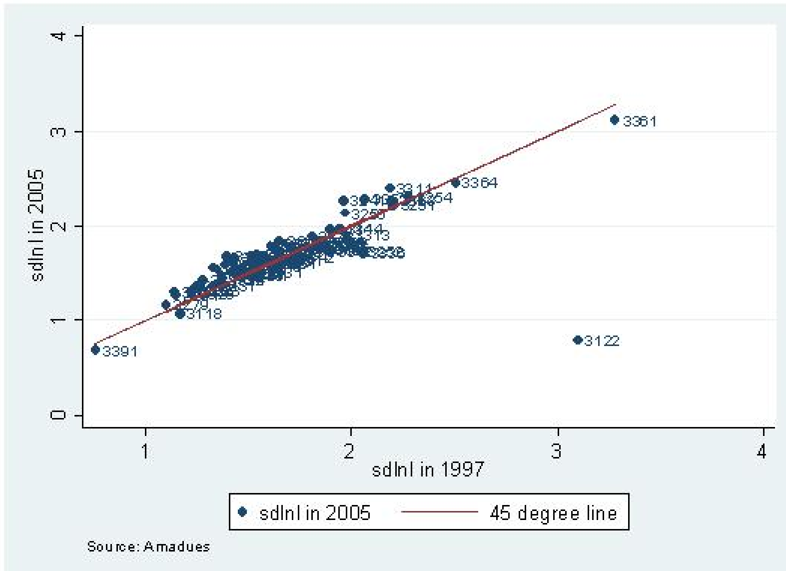


Figure 1.14: Sectoral Firm Size Heterogeneity in Chile 1979 and 1996

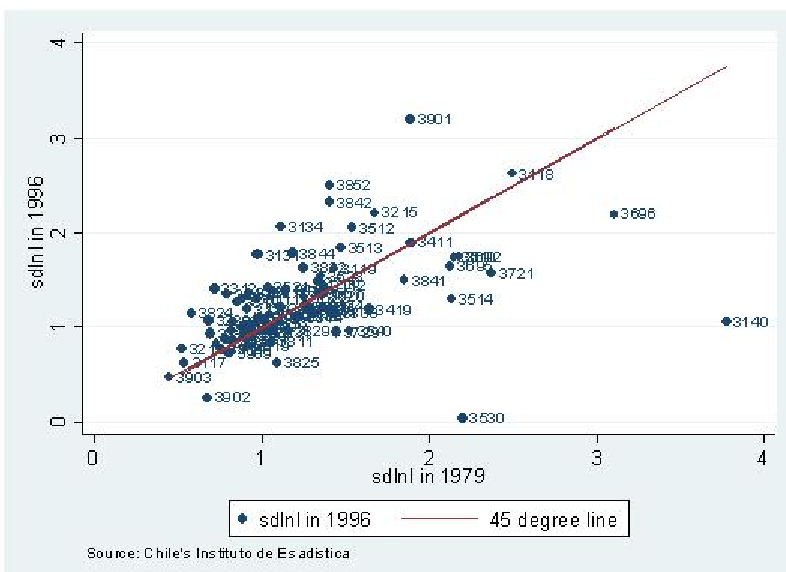


Figure 1.15: Sectoral Firm Size Heterogeneity in U.S. and France 1997

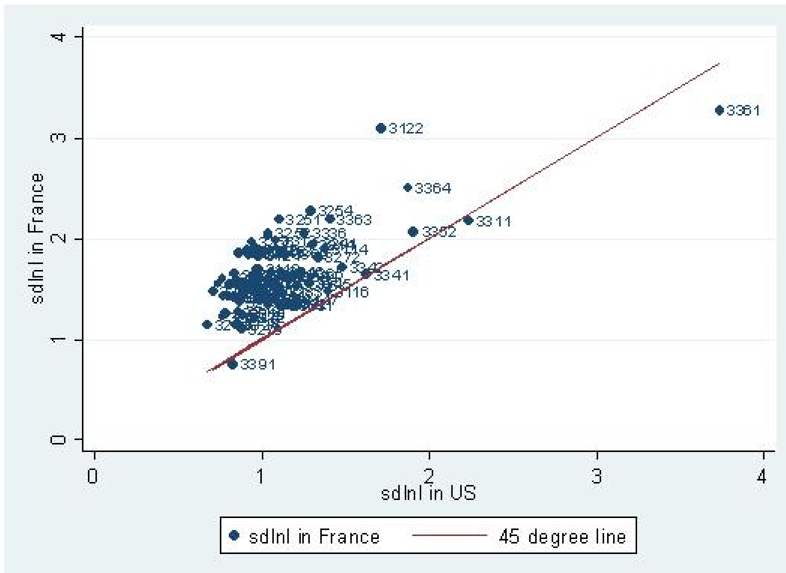


Figure 1.16: Sectoral Firm Size Heterogeneity in U.S. and France 2002

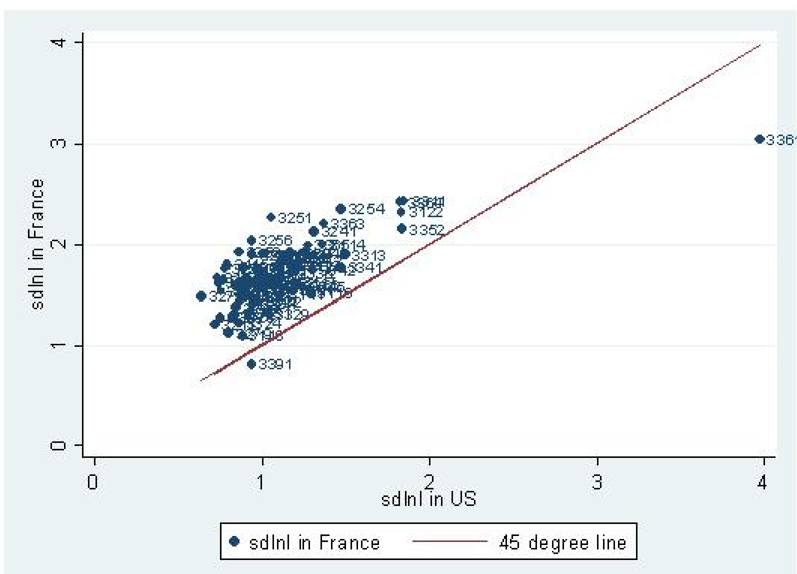


Figure 1.17: Simulated Cross-Firm Citation's Share in Total Citation by Sectors

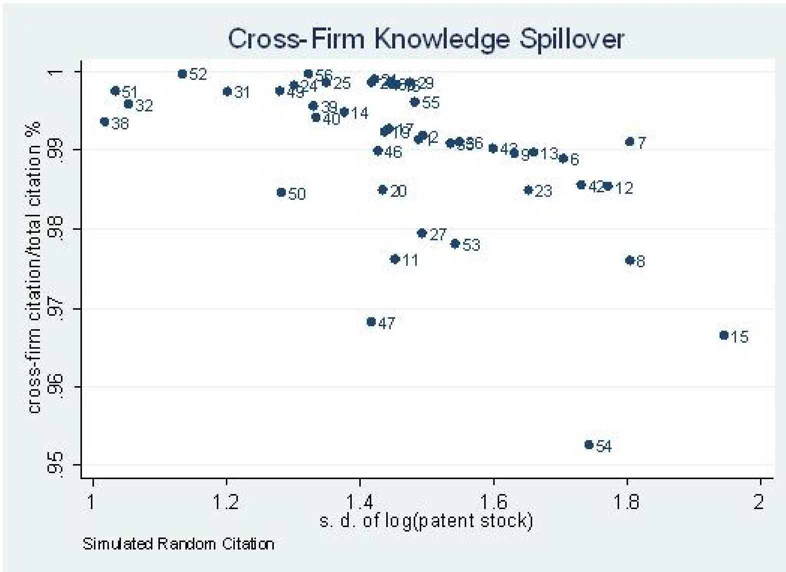


Figure 1.18: Scale-Independency of Innovation Rate by Sectors

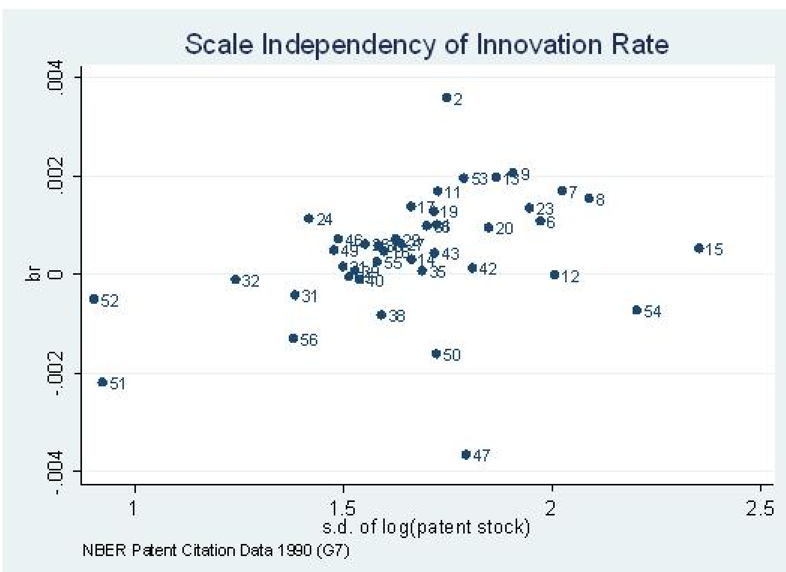


Figure 1.19: Scale-Dependency of Imitation Rate by Sectors

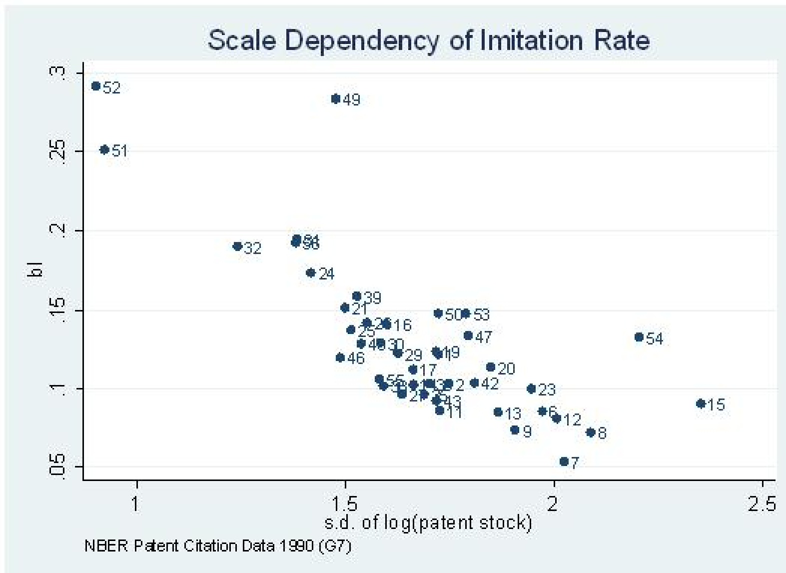


Figure 1.20: Determinant of Firm Size Heterogeneity

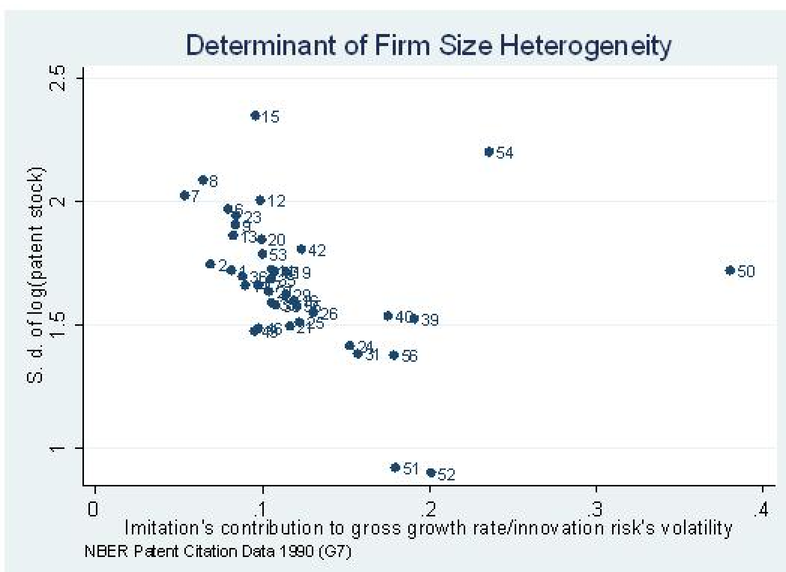
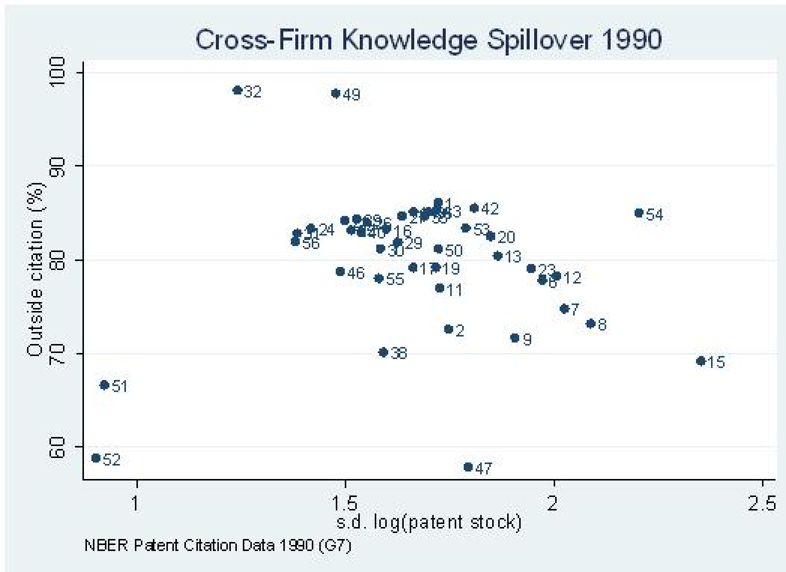


Figure 1.21: Cross-Firm Citation's Share in Total Citation by Sectors G7 Countries



Chapter 2

Dynamic Formation of Directed Networks

2.1 Introduction

The literature on dynamic networks formation among firms has primarily been concerned with understanding the formation of non-directed networks. It has been successful in explaining many empirical characteristics observed in actual social networks[1]. However, in reality, many networks transferring goods and information flows are directed. Examples of such networks are websites connected by hyperlinks, people connected by phone calls or emails, and firms connected by patent citations. In such directed networks, two nodes connected by one link are not symmetric but play different roles: initiator and receiver. People who send emails are not symmetric with those receiving them; patent citers are not the same as those being cited. In fact, for most networks, the symmetry of nodes is at best a

[1]Jackson and Rogers (38) summarize five characteristics: (1) small average shortest distance between nodes; (2) positive clustering coefficients (clustering coefficients measure how often two nodes with a common friend are also friends); (3) power-law degree distribution. A quantity x obeys a Power-law if it is drawn from a probability distribution $p(x) \propto x^{-\alpha}$, where α is a constant parameter of the distribution known as the exponent or scaling parameter. In real-world situations the scaling parameter typically lays in the range $2 < \alpha < 3$, although there are occasional exceptions; (4) positive correlation between degrees of linked nodes; and (5) negative correlation between the local clustering coefficient of a node's neighborhood and the node's degree.

simplifying assumption. There has, until now, been no model of dynamic networks formation for directed network. Any such model would need to provide insights into what leads selfish individuals to engage in network building, and to match the observed triple Power-law degree distributions of networks, i.e., the in-degree, out-degree and total-degree all follow Power-law degree distribution.

This paper builds a dynamic model of directed network formation and demonstrates its application to a real-world directed network: a firm citation network. The model uses profit sharing between a firm with access to a customer and a firm that possesses the technology to produce what the customer wants, to explain individual firms' incentives to build directed networks. When firms and customers are randomly matched, a representative firm may not be able to produce what its customer wants, but can gain by referring this customer to another firm whose production capability can meet the demand. The preconditions are that firms know what others in the networks produce and that they can profit, perhaps via commission fees, by directing consumers towards them. Firm i knows what firm j produces corresponds to a directed link pointing from firm i to firm j in the networks. With profit sharing as the form of commission, firms want to know other firms, so that they earn more commission fees, while at the same time they seek to become known by more other firms in order to obtain re-directed customers. In equilibrium, the dynamic process of network formation is decided by a firm's trade-off between the benefit and cost of building networks. Joharia, Mannorb, and Tsitsiklis (41) uses cost compensation to explain why private post offices build directed networks to deliver mail packages, but their model is static.

In a similar manner, we can understand a firm's incentive to build a directed knowledge network: to discover a new good, a firm needs complementary knowledge from two random fields. In most cases, one firm only masters the knowledge in one field and needs to learn the other field's knowledge from their peers. Since future knowledge demand is uncertain, each firm wants to link with more knowledge providers. But why are firms willing to provide knowledge to others without being paid directly? The key is the long term collaboration feature of the networks and the uncertainty of future knowledge demand. A teacher today has equal chance to learn from her/his learner in the future. Moreover, through communication with the learner, the teacher has a chance to build new links with unknown firms introduced by the learner. Therefore, the teacher is indirectly compensated by the chance to learn from more peers in the future. In other environments, as long as firm's economic activities need two complementary resources and each firm only owns one of them, a network that links the owners of both resources is a long term solution.

This paper extends the dynamic network formation of non-directed networks in Holme and Kim (34), Vazquez (69), and Jackson and Rogers (38) (JR henceforth) into directed networks[2]. In non-directed networks, nodes build a new link either through the network-based method (knowing a friend's friend) or randomized method (knowing random unknown people). In directed networks, there is another layer of complexity within the network-based network formation method: the two directions of links. Nodes build new links via both old links in the same direction and old links in the opposite direction at different success rates. The first success rate depends on a node's tendency to maintain its current role as an initiator or receiver in the networks; the second success rate relies on a node's possibility of switching its current role to the opposite role in the networks. According to Kesten (44), when the success rates in building new links through different methods are subject to i.i.d. 'popularity' shocks, the in-degree, out-degree, and total-degree distributions all converge to Power-law distribution, as seen in the real directed networks. The inter-temporal causality between two types of links in the bilateral networks-based networks formation is the key to generating the triple Power-law distribution.

The model can be extended to understand the dynamic formation of more complex networks, where there are multiple types of nodes and links; for example, an exporter-market network and a buyer-seller network. The key to handling complex networks is the modeling of the inter-temporal causality between different types of links.

I illustrate the application of the model by considering a firm citation network panel data. I construct the networks data from the National Bureau of Economic Research (NBER) Patent Citation Database for 42 sectors from 1985 to 1994. This

[2]Network-based network formation means two unconnected nodes with common neighbor in last period connect with each other the current period or 'knowing a friend's friend'). Its opposite is randomized network formation, where two randomly picked nodes connect with each other.

is clearly a directed network: one inter-firm citation corresponds to one directed link from the citing firm[3] to the cited firm[4]. With multiple years of data, I can observe the inter-temporal change in each sectoral citation networks. I can determine both the links that are newly built and the method by which the new links are built; whether a new link is introduced by an old link in the same direction, introduced by an old link in the opposite direction, or by the random meeting of two previously unconnected nodes. Identifying the method through which a node builds new connections allows me to infer the success rates of building new links via the different methods for each node. Knowing the distribution of these success rates, I am able to simulate the dynamic network formation process and compare the simulated networks with the real sectoral citation networks.

The simulation allows me to test whether the simple model of degree[5] dynamic process also mimics other structural features in the real networks. The simulated network for sector s starts from a randomly generated network. In each period, new links are built by a mixture of networks-based and randomized networks formation methods. In the bilateral network-based network formation, every node introduces its unconnected friends to each other. A representative node i at time t is assigned an i.i.d. 'popularity draw' po_{it}^{xy} from the estimated success rate distribution of building x type new link from y type old link in sector s, where x, y $\in$ (inward, outward). A higher popularity draw po_{it}^{xy} means a y type old link is more likely to introduce an x type new link to node i. In the randomized networks formation, two unconnected nodes i and j are randomly connected by a link from i to j by possibility r_s, which is the estimated success rate of random matching in sector s. Repeat the above process for 50 periods, which is long enough for degree distributions to converge to Power-law. The simulated networks match actual networks not only in degree distributions, but also in clustering coefficients and other structure features.

The remainder of the paper is organized as follows. The model section describes a firm's motivation to build a directed social network and the methods of building new links. The data section introduces the NBER Patent Citation Data

[3]The firm that applies a patent and cites other existing patents.

[4]A firm whose existing patents are cited by other patent applications.

[5]A node's degree is the number of nodes with which it directly connects.

and illustrates how to infer the distribution of popularity draws in each network formation method. With the estimated distribution of popularity draws, I simulate artificial networks and compare them with the real sectoral citation networks. Lastly, the conclusion summarizes the paper.

2.2 The Model

2.2.1 Differentiated Goods Market

There are N firms evenly distributed on $[0,1]$. Each firm plays two roles: producer and dealer. As a producer, firm i produces one unit of goods $i \in \left[i - \frac{1}{2N}, i + \frac{1}{2N}\right]$ with one unit of labor. A consumer wants a random good on $[0,1]$ each period. A customer for good j is willing to pay $P(1 - T|i-j|)$ dollars for good i. T measures the consumer's intolerance to product difference. Assume $T > 2N$, so that consumers only accept the closest substitute available.

As a dealer, firm i may receive a query from a randomly matched customer j for goods $j \in [0,1]$. If $j \in \left[i - \frac{1}{2N}, i + \frac{1}{2N}\right]$, firm i serves customer j itself. Otherwise, firm i acts as a dealer and checks other producers on its contact list $C_{it} = \{k | \ i \text{ knows } k\}$ at time t. If there exists one producer such that $k \in C_{it}$ and $j \in \left[k - \frac{1}{2N}, k + \frac{1}{2N}\right]$, dealer i introduces this customer to producer k and earns a commission fee $\theta(P - W)$. Producer k earns $(1-\theta)(P-W)$. The wage rate W is normalized to 1. θ represents a dealer's bargaining power. The market size is L.

More connections in the producer-dealer networks bring a higher expected profit to the firm. Denote the number of producers that i knows at time t (number of elements in C_{it}) as p_{it}. Denote the number of dealers that know i at time t (the number of firms k that $i \in C_{kt}$) as d_{it}. At time t, the expected profit for firm i that knows p_{it} producers and with d_{it} dealers knowing i is:

$$\pi_{it} = \frac{L}{N}\left(1 - \frac{T}{4N}\right)(P-1)\left[1 + \theta p_{it} + (1-\theta) d_{it}\right]. \tag{2.1}$$

The first part of the profit occurs when firm i serves the random customer itself. The second part originates from the commission fee, when firm i cannot serve the customer itself, but one of its producers can. The third part of the profit derives from the business introduced by d_{it} dealers that know firm i, when the dealers cannot serve their customers themselves. When there is a greater number of firms, each firm acquires a smaller market share $\frac{L}{N}$, but every realized trade brings higher profit $(1-\frac{T}{4N})$. This is because firms are more specialized and consumers will pay a higher price for a product tailored to their specific taste.

This dealer-producer relationship connects all firms by directed links: there is one directed link from firm i to its producer k. In networks theory, p_{it} is called the out-degree of firm i and d_{it} is called the in-degree of firm i. Firm i's profit (2.1) linearly increases with p_{it} and d_{it}.

2.2.2 Methods of Building Networks

According to the degree-dependent income flow in (2.1), any profit maximizing firm wants to expand its connections with other firms. As a dealer, there are three ways for firm i to know more producers: (1a) knowing a new producer k through a current producer j ($j \in C_{it}$, $k \in C_{jt}$ and $k \notin C_{it}$); (2a) knowing a new producer k through a current dealer j ($i \in C_{jt}$ and $k \in C_{jt}$); and (3a) randomly encountering a producer's advertisement on the street.

As a producer, there are three ways for firm i to acquire new dealers: (1b) ask a current dealer j to forward firm i's information to j's producer k ($i \in C_{jt}$, $k \in C_{jt}$ and $i \notin C_{kt}$) ; (2b) ask a current producer j to forward firm i's information to a firm j's producer k($j \in C_{it}$ and $k \in C_{jt}$); and (3b) send its advertisement to a random firm on the street.

Methods (1a), (2a), (1b), and (2b) belong to network-based networks formation, where firm i's new connections built today depend on its position in yesterday's networks. Methods (3a) and (3b) are called randomized networks formation, in which the new connections built today are independent of yesterday's networks topology.

2.2.3 Dynamic Networks Formation Process

Communication Technology

To communicate with current producers and dealers, firm i spends l_{it}^{p} hours listening and t_{it}^{p} hours talking to a producer; l_{it}^{d} hours listening and t_{it}^{d} hours talking to a dealer; and l_{it}^{r} hours listening and t_{it}^{r} hours talking to a random unknown firm. While listening, firm i receives information about more producers from the talkers; while talking, firm i broadcasts its own or another firm's information to the listeners. Denote the number of new y gained by communication with current x as $\triangle y_{ijt}^{x}$, in which $x \in \{$producer, dealer, and random firm$\}$, $y \in \{$producer, dealer$\}$, i is the talker, and j is the listener. For example, $\triangle p_{ijt}^{p}$ represents firm i's number of new producers known by listening to producer j ($j \in C_{it}$) at time t.

Since the communication is bilateral, the outcome depends on the time inputs from both sides. A higher number of current links means firm i is more experienced in communication, which helps it make more links in the future. Suppose firm j is firm i's producer, firm k is firm i's dealer;[6] firm r is a random unknown firm to firm i. The expected numbers of new links built through different methods are as follows:

$$\triangle p_{jit}^{p} = A_{pp}\left(t_{jt}^{d}\right)^{\beta}\left(l_{it}^{p}\right)^{\alpha}$$

$$\triangle d_{ijt}^{p} = A_{dp}\left(t_{it}^{p}\right)^{\beta}\left(l_{jt}^{d}\right)^{\alpha}$$

$$\triangle p_{kit}^{d} = A_{pd}\left(t_{kt}^{p}\right)^{\beta}\left(l_{it}^{d}\right)^{\alpha}$$

$$\triangle d_{ikt}^{d} = A_{dd}\left(t_{it}^{d}\right)^{\beta}\left(l_{kt}^{p}\right)^{\alpha}$$

$$\triangle p_{ilt}^{r} = A_{pr}\left(t_{rt}^{r}\right)^{\beta}\left(l_{it}^{r}\right)^{\alpha}$$

$$\triangle d_{ilt}^{r} = A_{dr}\left(t_{it}^{r}\right)^{\beta}\left(l_{rt}^{r}\right)^{\alpha}$$

Here $0 < \alpha < 1$, $0 < \beta < 1$, and $0 < \alpha + \beta < 1$. A_{xy} is the technology of knowing new x through current y, in which $y \in \{$producer, dealer, and random unknown firm$\}$, $x \in \{$producer and dealer$\}$. α and β measures the listener's and

[6] Notice that meanwhile firm i is also firm j's dealer and firm i is firm k's producer.

talker's share in the communication outcome, respectively. When $x, y \in \{$producer, dealer$\}$ A_{xy} measures the likelihood of trusting and building a link with an introduced firm. A_{xr} measures the likelihood of trusting and building links with random unknown firms.

Institutions and social norms affect the efficiency to build network by different methods directly and determine the networks structure indirectly. For example, assume government publishes each firm's credit history, so that any firm can check an unknown firm's record at a cost ϕ. A smaller ϕ makes it easier to trust an unknown firm, which corresponds to a higher A_{xr}. In a society where people discriminate against others of different backgrounds, firm owners may be very selective when building linkages with other firms. They may prefer to connect with others of similar background, and may trust an introduced firm better than a random firm. This type of social norm increases the productivity of network-based methods A_{xy} $x, y \in \{$producer, dealer$\}$ and lowers the productivity of randomized methods A_{xr}.

Firm's Problem

The firm i maximizes firm value by choosing hour inputs l_{it}^{p}, t_{it}^{p}, l_{it}^{d}, t_{it}^{d}, l_{it}^{r}, and t_{it}^{r}.

$$
\begin{aligned}
V_t\left(p_{it}, d_{it}\right) = \max \frac{L}{N}\left(1-\frac{1}{4N}\right)\left[P+\theta P p_{it}+\left((1-\theta)P-1\right)d_{it}\right] \\
+\rho E\left[V_{it+1}\left(p_{it+1}, d_{it+1}\right)\right]-p_{it}\left(l_{it}^{p}+t_{it}^{p}\right)-d_{it}\left(l_{it}^{d}+t_{it}^{d}\right)-Nl_{it}^{r}-Nt_{it}^{r}
\end{aligned}
$$

such that

$$
p_{it+1} = (1-\delta)\,p_{it}+\sum_{j\in C_i} A_{pp}\left(t_{jt}^{d}\right)^{\beta}\left(l_{it}^{p}\right)^{\alpha}+\varepsilon_{it}^{pp}p_{it}+ \tag{2.2}
$$

$$
\sum_{i\in C_k} A_{pd}\left(t_{kt}^{p}\right)^{\beta}\left(l_{it}^{d}\right)^{\alpha}+\varepsilon_{it}^{pd}d_{it}+\sum_{r\notin C_i} A_{pr}\left(t_{rt}^{r}\right)^{\beta}\left(l_{it}^{r}\right)^{\alpha}+\varepsilon_{it}^{pr}N_t \tag{2.3}
$$

$$d_{it+1} = (1-\delta)\, d_{it} + \sum_{j\in C_i} A_{dp} \left(t_{it}^{p}\right)^{\beta} \left(l_{jt}^{d}\right)^{\alpha} + \varepsilon_{it}^{dp} p_{it} + \tag{2.4}$$

$$\sum_{i\in C_k} A_{dd} \left(t_{it}^{d}\right)^{\beta} \left(l_{kt}^{p}\right)^{\alpha} + \varepsilon_{it}^{dd} d_{it} + \sum_{r\notin C_i} A_{pr} \left(t_{it}^{r}\right)^{\beta} \left(l_{rt}^{r}\right)^{\alpha} + \varepsilon_{it}^{dr} N_t \tag{2.5}$$

Firm value $V_t\,(p_{it}, d_{it})$ is a function of its in-degree and out-degree. (2.2) and (2.4) are the networks formation functions of out-degree and in-degree. ρ is the firm's discount rate. δ is the depreciation rate of contact information. In the general equilibrium, δ is set to the average speed at which dealers get to know new producers, so that the average number of producers per dealer is a constant overtime. Otherwise, when δ is too small, the network degenerates to a fully connected network; or when δ is too big, the network breaks down. $\left\{\varepsilon_{it}^{pp}, \varepsilon_{it}^{pd}, \varepsilon_{it}^{pr}, \varepsilon_{it}^{dp}, \varepsilon_{it}^{dd}, \text{and } \varepsilon_{it}^{dr}\right\}$ are the i.i.d. zero mean shocks that firm i receives in different types of social activities at time t. They capture firm i's random popularity in different types of social communication.

Firm i pays its representatives to gain more connections. In the first-order conditions (2.6) to (2.11), the expected marginal profit equals the marginal cost. Every firm takes other firms' time input as given.

$$l_{it}^{p} = (\rho V_p \alpha A_{pp})^{\frac{1}{1-\alpha}} \left(t_{jt}^{d}\right)^{\frac{\beta}{1-\alpha}} \tag{2.6}$$

$$l_{it}^{d} = \left(\rho V_p \alpha A_{pd}\right)^{\frac{1}{1-\alpha}} \left(t_{kt}^{p}\right)^{\frac{\beta}{1-\alpha}} \tag{2.7}$$

$$l_{it}^{r} = (\rho V_p \alpha A_{pr} N_t)^{\frac{1}{1-\alpha}} \left(t_{rt}^{r}\right)^{\frac{\beta}{1-\alpha}} \tag{2.8}$$

$$t_{it}^{p} = \left(\rho V_d \beta A_{dp}\right)^{\frac{1}{1-\beta}} \left(l_{kt}^{d}\right)^{\frac{\alpha}{1-\beta}} \tag{2.9}$$

$$t_{it}^{d} = (\rho V_d \beta A_{dd})^{\frac{1}{1-\beta}} \left(l_{kt}^{p}\right)^{\frac{\alpha}{1-\beta}} \tag{2.10}$$

$$t_{it}^{r} = (\rho V_d \beta A_{dr} N_t)^{\frac{1}{1-\beta}} \left(l_{rt}^{r}\right)^{\frac{\alpha}{1-\beta}} \tag{2.11}$$

An educated guess for the firm value function is

$$V\,(p,d) = v_p p + v_d d + u.$$

$$v_p = \frac{L}{N}\left(1-\frac{1}{4N}\right)(P-1)\,\theta \tag{2.12}$$

$$+\rho v_p\left(1-\delta+A_{pp}\left(t_{jt}^{d}\right)^{\beta}\left(l_{it}^{p}\right)^{\alpha}\right)+\rho v_d A_{pp}\left(t_{jt}^{d}\right)^{\beta}\left(l_{it}^{p}\right)^{\alpha}-\left(l_{it}^{p}+t_{it}^{p}\right) \tag{2.13}$$

$$v_d = \frac{L}{N}\left(1-\frac{1}{4N}\right)(P-1)(1-\theta) \tag{2.14}$$

$$+\rho v_p A_{pd}\left(t_{kt}^{p}\right)^{\beta}\left(l_{it}^{d}\right)^{\alpha}+\rho v_d\left(1-\delta+A_{dp}\left(t_{it}^{p}\right)^{\beta}\left(l_{jt}^{d}\right)^{\alpha}\right)-\left(l_{it}^{d}+t_{it}^{d}\right) \tag{2.15}$$

$$u = \frac{L}{N}\left(1-\frac{1}{4N}\right)P+\rho v_p A_{pr}\left(t_{rt}^{r}\right)^{\beta}\left(l_{it}^{r}\right)^{\alpha}+\rho v_d A_{pr}\left(t_{it}^{r}\right)^{\beta}\left(l_{rt}^{r}\right)^{\alpha}-Nl_{it}^{r}-Nt_{it}^{r}$$

To firm i, one link's marginal value is equal to the discounted value of future profit from it and new links introduced by that link, minus the cost of communiction with linked firms. With free entry, the number of firms reaches the equilibrium when u is equal to the fixed entry cost f_e. For a potential entrant that has no linkage with any incumbent, its firm value includes only u, which is equal to the profit from exact matched consumers plus the profit from links made with random firms, minus the cost of communication with random firms.

First order conditions (2.6) to (2.11) show that firm i exerts more effort in knowing new producers (dealers), when its marginal value v_p (v_d) is higher. In general equilibrium, every firm chooses the same time input portfolio $\{l_{it}^{p*}$, t_{it}^{d*}, l_{it}^{d*}, t_{it}^{p*}, l_{it}^{r*}, and $t_{it}^{r*}\}$.

$$l_{it}^{p*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_p\alpha A_{pp}\right)^{\frac{1-\beta}{1-\alpha-\beta}}\left(v_d\beta A_{dd}\right)^{\frac{\beta}{1-\alpha-\beta}} \tag{2.16}$$

$$t_{it}^{d*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_d\beta A_{dd}\right)^{\frac{1-\alpha}{1-\alpha-\beta}}\left(v_p\alpha A_{pp}\right)^{\frac{\alpha}{1-\alpha-\beta}} \tag{2.17}$$

$$l_{it}^{d*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_p\alpha A_{pd}\right)^{\frac{1-\beta}{1-\alpha-\beta}}\left(v_d\beta A_{dp}\right)^{\frac{\beta}{1-\alpha-\beta}} \tag{2.18}$$

$$t_{it}^{p*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_d\beta A_{dp}\right)^{\frac{1-\alpha}{1-\alpha-\beta}}\left(v_p\alpha A_{pd}\right)^{\frac{\alpha}{1-\alpha-\beta}} \tag{2.19}$$

$$l_{it}^{r*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_p\alpha A_{pr}\right)^{\frac{1-\beta}{1-\alpha-\beta}}\left(v_d\beta A_{dr}\right)^{\frac{\beta}{1-\alpha-\beta}} \tag{2.20}$$

$$t_{it}^{r*} = \rho^{\frac{1}{1-\alpha-\beta}}\left(v_d\beta A_{dr}\right)^{\frac{1-\alpha}{1-\alpha-\beta}}\left(v_p\alpha A_{pr}\right)^{\frac{\alpha}{1-\alpha-\beta}} \tag{2.21}$$

$$v_p\left(1-\rho\left(1-\delta\right)\right)=\frac{L}{N}\left(1-\frac{1}{4N}\right)P\theta+\left(\frac{1}{\alpha}-1\right)l_{it}^{p*}+\left(\frac{1}{\beta}-1\right)t_{it}^{p*}$$

$$v_d\left(1-\rho\left(1-\delta\right)\right)=\frac{L}{N}\left(1-\frac{1}{4N}\right)\left(P\left(1-\theta\right)-1\right)+\left(\frac{1}{\alpha}-1\right)l_{it}^{d*}+\left(\frac{1}{\beta}-1\right)t_{it}^{d*}$$

$$f_e=u=\frac{L}{N}\left(1-\frac{1}{4N}\right)P+N\left(\frac{1}{\alpha}-1\right)l_{it}^{r*}+N\left(\frac{1}{\beta}-1\right)t_{it}^{r*}$$

The equilibrium communication inputs l_{it}^{p*}, t_{it}^{d*}, l_{it}^{d*}, t_{it}^{p*}, l_{it}^{r*}, and t_{it}^{r*}, firm value function parameters v_p and v_d and the number of firms N are jointly determined by the 9 equations above.

There is another way to explain the second part of v_p and v_d, $\left(\frac{1}{\alpha}-1\right)l_{it}^{p*}+\left(\frac{1}{\beta}-1\right)t_{it}^{p*}$ and $\left(\frac{1}{\alpha}-1\right)l_{it}^{d*}+\left(\frac{1}{\beta}-1\right)t_{it}^{d*}$. They capture the externality from other linked firms. The reason is that building new links depends on the efforts of both the listener and the talker, but each party only pays for its own effort. One party's higher input also induces a larger input from the other party. As a result, in a network where every firm inputs more time communicating with others, the value of each link is higher due to the bilateral externality.

Notice that v_p is still positive even if a producer does not share profit with a dealer when $\theta=0$, because the dealer is rewarded with the externality from the producer's communication effort, which is the chance to know new dealers and producers while talking and listening to the producer. In the firm citation networks, a dealer acts as a knowledge teacher, who provides the type of knowledge that is complementary to the producer's knowledge. θ measures the strength of intellectual property rights protection (IPRP). Under the worse institution of IPRP ($\theta=0$), although firms' communication efforts are not maximized, the knowledge networks is still sustainable because the teacher is rewarded with the possibility of knowing new teachers and learners in the future.

Nevertheless, an extremely strong IPRP or θ approaches 1 is not optimal either. In a special case where $\alpha=\beta$ and matrix A is symmetric, firms exert the largest communication efforts when $\theta=0.5$ and $v_p=v_d$. That is due to the complementaries between the efforts from two parties: when θ deviates from 0.5, the party that gains a smaller share of profit reduces its communication input; the other party also reduces its effort because it enjoys a smaller externality.

Substituting Nash equilibrium time input (2.16) to (2.21) into the networks formation functions (2.4) and (2.2) solves the dynamic networks formation processes for firm i.

$$\begin{pmatrix} p_{it+1} \\ d_{it+1} \end{pmatrix} = F_t \begin{pmatrix} p_{it} \\ d_{it} \end{pmatrix} + R_t \tag{2.22}$$

$$F_t = \begin{pmatrix} F^{pp} & F^{pd} \\ F^{dp} & F^{dd} \end{pmatrix}$$

$$F_t^{pp} = 1 - \delta + (v_p \alpha A_{pp})^{\frac{\alpha}{1-\alpha-\beta}} (v_d \beta A_{dd})^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{pp}$$

$$F_t^{pd} = \left(v_p \alpha A_{pd}\right)^{\frac{\alpha}{1-\alpha-\beta}} \left(v_d \beta A_{dp}\right)^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{pd}$$

$$F_t^{dp} = \left(v_p \alpha A_{pd}\right)^{\frac{\alpha}{1-\alpha-\beta}} \left(v_d \beta A_{dp}\right)^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{dp}$$

$$F_t^{dd} = 1 - \delta + (v_p \alpha A_{pp})^{\frac{\alpha}{1-\alpha-\beta}} (v_d \beta A_{dd})^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{dd}$$

$$R_t = \begin{pmatrix} (v_p \alpha A_{pr})^{\frac{\alpha}{1-\alpha-\beta}} (v_d \beta A_{dr})^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{pr} \\ (v_p \alpha A_{pr})^{\frac{\alpha}{1-\alpha-\beta}} (v_d \beta A_{dr})^{\frac{\beta}{1-\alpha-\beta}} + \varepsilon_{it}^{dr} \end{pmatrix}$$

Notice that the expectation of matrix F_t is center-symmetric, because every firm's inputs are the same. For example, firm i spends as much time listening to its producer j as its dealer k listens to firm i. Firm j also spends as much time talking to firm i as firm i talks to firm k, which is why the upper-left and lower-right elements of $E(F)$ are the same. Similarly, the upper and lower elements of $E(R)$ are symmetric, because firm i spends the same time listening (talking) to a random firm as a random firm spends listening (talking) to firm i.

Proposition 4 *According to Kesten (44), when* $\{\varepsilon_{it}^{pp}, \varepsilon_{it}^{pd}, \varepsilon_{it}^{pr}, \varepsilon_{it}^{dp}, \varepsilon_{it}^{dd}$, *and* $\varepsilon_{it}^{dr}\}$ *are identically independent distributed across firm and time, for two-dimensional vector x with* $|x| = 1$, *as* $t \to \infty$, $x' \binom{p_{it}}{d_{it}}$ *follows Pareto distribution* μ_x.

By choosing $x = (1,0),(0,1)$, and $(\frac{1}{\sqrt{2}}, \frac{1}{\sqrt{2}})$, I obtain the distributions of out-degree (p_{it}) in-degree (d_{it}) and total degree ($p_{it} + d_{it}$) of the networks. Since the matrices F and R in (2.22) are symmetric, p_{it} and d_{it} have the same *Pareto* distribution parameter μ. Notice that Power-law distribution is the discrete time version of *Pareto* distribution. Since the number of links is a discrete number, the out-degree in-degree and total degree exhibit Power-law distributions.

2.3 Welfare

In this simple environment, network density, the number of firms (degree of specialization), and the heterogeneity of degree distribution jointly determine consumer welfare. Communication technologies, fixed market entry cost, IPRP institutions and social norms influencing firm's attitudes to unknown firms and introduced firms are the fundamentals of setting the above networks features.

2.3.1 Risk-Neutral Consumers

The social welfare $L\frac{E(d)}{N}\left(1-\frac{T}{4N}\right)$ depends on the network density[7] $\frac{E(d)}{N}$ and the degree of specialization $\frac{1}{N}$. Higher network density improves the chance of realizing a trade, while specialization increases consumer utility from each trade. For networks with a constant networks density, the contact depreciation rate δ must be equal to the rate at which new links are built.

$$\delta = A_{pp}\left(t_{jt}^{d}\right)^{\beta}\left(l_{it}^{p}\right)^{\alpha} + A_{pd}\left(t_{kt}^{p}\right)^{\beta}\left(l_{it}^{d}\right)^{\alpha} + \frac{N}{E\left(d\right)}A_{pr}\left(t_{rt}^{r}\right)^{\beta}\left(l_{it}^{r}\right)^{\alpha}$$

Therefore, network density decreases with the depreciation rate of links δ and increases with the productivity to build links via different methods A_{pp}, A_{pd} and A_{pr}. In a knowledge network, an IPRP that divides profit equally between learner and teacher maximizes the communication inputs from both sides; in the meantime, network density is also maximized. As long as network density is a constant, more firms in the industry leads to deepening specialization without shrinking each firm's market share or likelihood of trade.

[7] The average number of outward (inward) links of each firm in the network.

2.3.2 Risk-Averse Consumers

When consumers are risk-averse in terms of trade likelihood, the variance of degree distribution enters the welfare function. When a consumer wants a random good j each period, a more heterogeneous degree distribution lowers the consumer's expected utility at a given network density, because the trade likelihood varies too greatly across types of goods.

Referring to the discussion in the 'Degree Distribution Heterogeneity and Ratio r' section below, heterogeneity of degree distribution decreases with the relative productivity between randomized method and network-based method to build new links. For example, if firms can easily trust a random unknown, A_{dr} is higher relative to A_{dd} and A_{dp}, therefore a higher percentage of links are built through the randomized method instead of network-based methods. When the randomized method dominates, all firms have a relatively equal number of links. In contrast, when network-based methods dominate, the well-connected firms today will be even better connected tomorrow. As richer become richer, the degree distribution becomes more dispersed. Referring to the discussion in the 'Communication Technology' section below, a higher cost ϕ to check an unknown firm's credit record or a discriminative social norm induce lower A_{dr} relative to A_{dd} and A_{dp}, hence a more heterogeneous degree distribution.

In summary, regardless of consumer risk attitudes, social welfare increases with the productivity to build and preserve linkages among firms and decreases with the fixed market entry cost. When consumers are risk-averse, institutions and social norms that encourage firms to trust and connect with random unknown firms lead to more homogeneous degree distribution and higher consumer welfare.

2.4 Data

2.4.1 Data Description

The NBER Patent Citation Database published by the U.S. Patent Office reports patent applications in 42 broad SIC classifications from 1962 to 2002. With multiple years, it allows me to track the inter-temporal change of networks. With 42 sectors, it is also convenient for comparing cross sectors. I use within-sector inter-firm citations made between U.S. firms from 1985 to 1995 to construct sectoral citation networks for 42 sectors.

Figure 2.1 shows the three-dimensional graph of a real firm citation networks based on the Refrigeration and Service Industry Machinery sector during the period 1990-1994. Each berry represents a firm, and a link with arrow indicates a citation from the citing firm to the cited firm. There are numerous layers in the networks: firms with more links lay closer to the center of the networks, while firms with fewer links stay in the periphery of the networks.

A sectoral citation network is constructed as follows. Every firm is a node in the networks. Every citation is a directed link pointing from the citing firm to the cited firm. At time t, an n by n adjacency matrix M_{st} summarizes sector s' citation networks, where n is the total number of firms. $M_{st}(i,j) = 1$ if firm i cites firm j; otherwise $M_{st}(i,j) = 0$. Firm i's out-degree (number of outward links or producers p_{it} in the model) is the number of ones in the i^{th} row of M_{st}. Firm i's in-degree (number of inward links or dealers d_{it} in the model) is the number of ones in the i^{th} column of $M_{s,t}$. The total number of inward and outward links is called total-degree. Denote it as $t_{it} = p_{it} + d_{it}$.

2.4.2 Stylized Facts about Citation networks

Newman (59) and JR summarize five stylized facts that socially generated networks share. Sectoral firm citation networks have all these characteristics. Notice that these five facts are all static, not dynamic, features of networks. I report (a), (d) and (e) in Table 1, and (b) and (c) in Table 2.2.

(a) Average shortest distance between pairs of nodes is small.

(b) As with the other social networks, clustering coefficients[8] are larger than those in randomly generated networks.

(c) Power-law in-degree (d), out-degree (p), and total-degree (t) distribution (triple Power-law).

(d) Positive sorting. Degrees and patent stocks of linked firms are positively correlated. Average geographic distances between the citing and the cited firms are much shorter than average distance between two randomly picked firms in the

[8] They measure how likely it is that two nodes with a common connected node are also connected.

same sector.

(e) The clustering among the neighbors of a given node is inversely related to the node's degree.

2.4.3 Degree Distribution Heterogeneity and Ratio r in JR

JR predicts that the networks degree distribution is more heterogeneous and networks structure is more clustered when nodes build more new links with a friend's friend, and form fewer new links with a random node. Because a connection with a friend's friend is a type of 'preferential attachment', it means nodes with more links today acquire more new links tomorrow. In contrast, every node has an equal chance of building new links with random nodes in the randomized networks formation, which tends to eliminate current differences in degree numbers. The patent citation data contains directed firm network information for more than 30 years and 42 sectors, which permits me to test these predictions across sectors.

For sector s at time t, I estimate the key parameter r_{st} in JR, the ratio of new links with a random node to new links with a friend's friend. On the other hand, I also estimate the Power-law degree distribution parameter μ_{st} and calculate three measures of clustering coefficient C_{st}^{TT}, C_{st}, and C_{st}^{Avg} listed in JR. I give the details of estimating r_{ts}, μ_{st}^{in}, μ_{st}^{out}, μ_{st}^{total}, C_{st}^{TT}, C_{st}, and C_{st}^{Avg} in the Appendix 2.1.

As predicted in their paper, in Figures 2.2, 2.2, and 2.2, μ_{st}^{in}, μ_{st}^{out}, and μ_{st}^{total} are higher (degree distribution is more homogeneous), when r_{ts} is higher in sector s. In Figure 2.5 to Figure 2.7, C_{st}^{TT}, C_{st}, and C_{st}^{Avg} are smaller, when r_{ts} is higher in sector s.

2.4.4 Simulation

To test whether the networks formation process specified in the model section mimics the real networks formation process, I simulate a directed network for every sector and compare the simulated networks with the real sectoral networks. Before simulating, I must estimate the distribution of random matrices F_s and R_s in (2.22) ,$G_{F_s}(F_s)$ and $G_{R_s}(R_s)$, for every sector s. I give the details for estimating $G_{F_s}(F_s)$ and $G_{R_s}(R_s)$ in Appendix B. I then fit $\{F_{ft}\}$ and $\{R_{ft}\}$ with log-normal distribution. δ_s is set to be the average growth rate of new links. The detailed simulation process is listed in Appendix C.

In Figure 2.8 to Figure 2.13, I compare the degree distribution parameters $\mu_{st}^{in}(t)$, $\mu_{st}^{out}(t)$, and $\mu_{st}^{total}(t)$ as well as the clustering coefficients $C_{st}^{TT}(t)$, $C_{st}(t)$, and $C_{st}^{Avg}(t)$ in the simulated networks with their value in real sectoral networks. Each dot represents a sector. The straight line is the 45 degree line. The $\mu_{st}^{in}(t)$, $\mu_{st}^{out}(t)$, $\mu_{st}^{total}(t)$, $C_{st}^{TT}(t)$, $C_{st}(t)$, and $C_{st}^{Avg}(t)$ reported for simulated networks are the average value of the last 20 periods. The values reported for real networks are the five-year average from 1991 to 1995.

The simulated networks mimic the real networks in terms of degree distributions and clustering coefficients ($\mu_{st}^{in}(t)$, $\mu_{st}^{out}(t)$, $\mu_{st}^{total}(t)$, and $C_{st}^{TT}(t)$. $C_{st}(t)$ and $C_{st}^{Avg}(t)$ in simulated networks deviate from their correspondents in real networks, but the rank across sectors is still retained. The more highly clustered sectors in the real world are still more highly clustered in the simulated world.

In Figure 2.14, I compare the simulated value and real value of the correlation between the clustering among the neighbors of a given node and the node's degree for all sectors. Although most sectors still exhibit negative correlation between the clustering among the neighbors of a given node and the node's degree, the simulated networks abandon the cross-sector rank among real networks.

Among the real networks, the more highly clustered sectors also have more heterogeneous degree distribution (lower $\mu_{st}^{in}(t)$, $\mu_{st}^{out}(t)$, and $\mu_{st}^{total}(t)$) as displayed in Figure 2.15. Figure 2.16 shows that this rank is also maintained in the simulated world.

In conclusion, the simulated sectoral networks have a structure similar to their corresponding real networks. The cross-sector ranks in many structure measures are preserved in all but one case.

2.5 Conclusion

This paper extends the current literature in dynamic networks formation to the directed networks. The model uses profit sharing to explain a firms' motivation to build directed networks. Firms with customer access may not have the technology to produce what the consumer wants. Within a directed network, the firm with customer access introduces the customer to the firm with the required technology and receives a commission fee as an incentive. Since the communication outcome depends on the efforts of both parties, there is an externality from staying in touch with linked firms. Moreover, a knowledge network is still sustainable even when the knowledge teacher is not rewarded by profit sharing; instead the reward comes from the opportunity to know new teachers and learners in the future through the communication with the current learner.

The model extends the network-based networks formation method in Jackson and Rogers (38) by modeling the inter-temporal causality between two types of links. A current link in one direction may introduce new links in both directions. The inter-temporal causality between links in two directions is the key to generating triple Power-law degree distribution of in-degree, out-degree and total-degree, as observed in real directed networks.

Networks allow firms to become more specialized without losing customers. The reason is that a higher number of firms in the market also bring more potential dealers who redirect customers. Social welfare increases with the efficiency to build and preserve linkages among firms, because of higher network density and trade likelihood. Due to the complementaries between communication efforts from two parties, an intellectual property rights institution that equally divides the profit between the knowledge owner and the learner maximizes the communication efforts from both parties and the network density. Lower fixed market entry cost also improves welfare by hosting a larger number of more specialized firms. When consumers are risk-averse, institutions and social norms that encourage firms to trust and connect with random unknown firms lead to a more homogeneous degree distribution, which increases consumer welfare by reducing the variance of cross-good trade likelihood.

The empirical part of the paper constructs the sectoral firm citation networks from the NBER Patent Citation Database and estimates the model parameters from the panel networks data. The simulated networks have a structure similar to their real counterparts. Meanwhile, the empirical section also proves the predictions in Jackson and Rogers (38). In future research, the extended model can be used to understand the dynamic formation of more complex networks with multiple types of nodes and links.

Table 2.1: Stylized Facts (a), (d) and (e) for Firm Citation Networks

1 Patent Category	2 SIC87 Code	3 Average Shortest Distance between Two Nodes (Links)	4 Correlation in Total Degree	5 Correlation in Patent Stock (Significant level)	6 Great Circle Distance between Random Nodes (Kilometer)	7 Great Circle Distance between Linked Nodes (Kilometer)	8 Correlation between Local Clustering and Total Degree
1	20	3.568	0.256 (0.00)	0.174 (0.00)	6941.487	1061.766	-0.059
2	22	4.161	0.335 (0.00)	0.274 (0.00)	6991.833	793.612	-0.057
6	281	4.347	0.223 (0.00)	0.259 (0.00)	6487.991	988.637	-0.052
7	286	4.897	0.334 (0.00)	0.309 (0.00)	7159.114	724.757	-0.062
8	282	3.833	0.262 (0.00)	0.219 (0.00)	6724.499	724.359	-0.048
9	287	4.065	0.326 (0.00)	0.317 (0.00)	6835.288	839.865	-0.059
11	284	3.311	0.318 (0.00)	0.285 (0.00)	6250.190	804.956	-0.051
12	285	4.130	0.189 (0.00)	0.222 (0.00)	6735.441	662.364	-0.048
13	289	5.133	0.374 (0.00)	0.216 (0.00)	6643.541	844.491	-0.056
14	283	3.888	0.198 (0.00)	0.143 (0.00)	6366.191	1498.119	-0.058
15	1329	3.249	0.175 (0.00)	0.276 (0.00)	6566.865	894.687	-0.051
16	30	3.311	0.250 (0.00)	0.367 (0.00)	6950.680	994.404	-0.057
17	32	5.817	0.336 (0.00)	0.428 (0.00)	6890.830	918.441	-0.061
19	331+	2.903	0.215 (0.00)	0.127 (0.00)	6534.431	606.400	-0.036
20	333+	3.484	0.267 (0.00)	0.197 (0.00)	6769.626	963.375	-0.021
21	34-	5.574	0.259 (0.00)	0.308 (0.00)	6766.991	1064.213	-0.058
23	351	3.580	0.159 (0.00)	0.370 (0.00)	6766.586	939.101	-0.063
24	352	5.963	0.270 (0.00)	0.285 (0.00)	7044.228	1033.380	-0.036
25	353	5.962	0.197 (0.00)	0.383 (0.00)	6777.036	918.523	0.000
26	354	5.424	0.180 (0.00)	0.296 (0.00)	6983.690	941.041	-0.048
27	357	3.580	0.167 (0.00)	0.191 (0.00)	6687.346	1618.286	-0.062
29	355	6.037	0.201 (0.00)	0.328 (0.00)	6648.464	891.189	-0.058
30	356	5.947	0.180 (0.00)	0.315 (0.00)	6695.607	929.471	-0.060
31	358	4.035	0.223 (0.00)	0.214 (0.00)	7215.740	1116.276	-0.025
32	359	1.675	0.294 (0.00)	0.243 (0.00)	6127.337	1262.372	-0.045
35	361+	3.626	0.105 (0.00)	0.211 (0.00)	6974.893	1305.647	-0.050
36	362	3.978	0.152 (0.00)	0.246 (0.00)	7042.136	1016.173	-0.057
38	363	2.123	0.191 (0.00)	0.289 (0.00)	6359.496	670.735	-0.017
39	364	4.759	0.285 (0.00)	0.378 (0.00)	6888.798	1174.515	-0.057
40	369	4.059	0.233 (0.00)	0.299 (0.00)	6863.550	1238.117	-0.055
42	365	3.315	0.144 (0.00)	0.183 (0.00)	7250.125	1236.902	-0.061
43	366+	4.280	0.170 (0.00)	0.251 (0.00)	6953.626	1325.210	-0.059
46	371	3.764	0.215 (0.00)	0.457 (0.00)	7060.468	944.429	-0.056
47	376	2.404	-0.061 (0.00)	0.249 (0.00)	7256.875	1975.178	0.212
49	373	1.967	0.268 (0.00)	0.102 (0.00)	6917.394	1721.915	0.056
50	374	3.891	0.115 (0.00)	0.273 (0.00)	5475.854	775.826	-0.045
51	375	2.238	0.194 (0.00)	0.199 (0.00)	6505.385	1770.709	0.191
52	379-	1.589	0.044 (0.00)	0.065 (0.00)	5743.242	1252.619	NaN
53	348+	4.113	0.257 (0.00)	0.271 (0.00)	6884.717	1379.926	-0.018
54	372	2.872	-0.002 (0.95)	0.336 (0.00)	5551.671	1193.121	-0.053
55	38-	6.257	0.193 (0.00)	0.412 (0.00)	6945.959	1379.452	-0.058
56	99	6.860	0.277 (0.00)	0.270 (0.00)	6904.848	1336.209	-0.054

Table 2.2: Stylized Facts (b) and (c) for Firm Citation Networks

1	2	3	4	5	6	7	8	9
Patent Category	SIC87 Code	Mu-in[9]	Mu-out	Mu-total	Mu-ps	CTT	C	Cavg
1	20	.666 (.97)	.592 (.50)	.639 (.88)	1.941 (.88)	.045	.043	.139
2	22	.817 (.91)	.724 (.10)	.791 (.38)	1.856 (.31)	.050	.049	.178
6	281	.744 (.93)	.675 (.73)	.717 (.90)	1.839 (.61)	.056	.054	.151
7	286	.791 (.78)	.714 (.62)	.756 (.97)	1.635 (0)	.055	.052	.229
8	282	.656 (.26)	.576 (.10)	.638 (.72)	1.663 (.02)	.056	.055	.173
9	287	.794 (.17)	.723 (.47)	.768 (.01)	1.584 (.01)	.055	.053	.235
11	284	.612 (.98)	.521 (.10)	.603 (.96)	1.843 (1.00)	.049	.047	.135
12	285	.876 (.64)	.753 (.01)	.825 (.84)	1.486 (.30)	.053	.052	.150
13	289	.692 (.92)	.653 (.59)	.680 (.95)	1.898 (.61)	.049	.048	.161
14	283	.925 (.74)	.905 (.27)	.910 (.49)	1.710 (.37)	.046	.046	.188
15	1329	.568 (.61)	.508 (.10)	.551 (.66)	1.713 (.087)	.041	.040	.122
16	30	.888 (.45)	.849 (.99)	.866 (.76)	1.803 (.24)	.054	.051	.243
17	32	.867 (.84)	.749 (.05)	.829 (.24)	2.087 (.95)	.050	.050	.174
19	331+	.885 (.56)	.720 (.10)	.837 (.65)	1.881 (.27)	.023	.022	.055
20	333+	.850 (.40)	.714 (.64)	.804 (.32)	2.263 (.83)	.038	.041	.085
21	34-	1.035 (.93)	.961 (.92)	1.003 (.69)	2.052 (.20)	.054	.052	.223
23	351	.567 (.32)	.539 (.51)	.571 (.31)	1.722 (.77)	.051	.050	.178
24	352	1.043 (.65)	.933 (.70)	1.010 (.89)	2.136 (.76)	.040	.039	.104
25	353	.907 (.97)	.852 (.93)	.891 (.77)	2.139 (.78)	.053	.049	.161
26	354	1.146 (.60)	1.010 (.10)	1.091 (.38)	1.844 (.58)	.042	.042	.133
27	357	.572 (.13)	.566 (.41)	.580 (.17)	1.731 (.60)	.056	.053	.298
29	355	.983 (.84)	.936 (.19)	.967 (.80)	1.697 (.65)	.054	.050	.198
30	356	.958 (.98)	.859 (.56)	.920 (.10)	1.787 (.20)	.053	.051	.234
31	358	1.204 (.51)	1.018 (.10)	1.155 (.03)	1.870 (.12)	.042	.039	.074
32	359	1.494 (.10)	1.021 (.10)	1.300 (.10)	1.937 (.95)	.008	.008	.038
35	361+	.826 (.94)	.803 (.85)	.817 (.87)	1.855 (.89)	.059	.057	.177
36	362	.896 (.45)	.836 (.84)	.890 (.54)	1.963 (.95)	.048	.048	.179
38	363	.780 (.52)	.566 (.10)	.744 (.01)	1.867 (.21)	.017	.012	.028
39	364	.953 (.15)	.858 (.38)	.907 (.89)	1.850 (.28)	.047	.044	.139
40	369	.750 (.97)	.659 (.22)	.722 (.74)	1.713 (.11)	.050	.050	.166
42	365	.727 (.95)	.709 (.10)	.731 (.41)	1.500 (.28)	.048	.045	.173
43	366+	.626 (.99)	.614 (.65)	.622 (.82)	1.759 (.68)	.056	.052	.295
46	371	.649 (.10)	.608 (.10)	.649 (.88)	1.760 (.93)	.055	.054	.172
47	376	.963 (.10)	.742 (.01)	.911 (.50)	1.524 (.52)	.010	.009	.008
49	373	1.259 (.10)	.915 (.10)	1.122 (.39)	1.906 (.57)	.015	.013	.022
50	374	.797 (.10)	.625 (.10)	.743 (.05)	2.170 (.44)	.018	.017	.063
51	375	1.540 (.01)	.986 (.10)	1.345 (.72)	1.660 (.01)	.012	.007	.014
52	379-	1.157 (.99)	.683 (.10)	1.005 (.442)	1.730 (.01)	.000	.000	.000
53	348+	.867 (.61)	.741 (.10)	.844 (.77)	1.801 (.50)	.034	.036	.069
54	372	.809 (.35)	.689 (.10)	.782 (.53)	1.567 (.17)	.015	.015	.059
55	38-	.681 (.50)	.696 (.93)	.692 (.97)	1.859 (.74)	.056	.051	.282
56	99	.931 (.74)	.869 (.32)	.911 (.25)	1.839 (.19)	.049	.047	.185

Figure 2.1: An Example of Firm Citation Networks

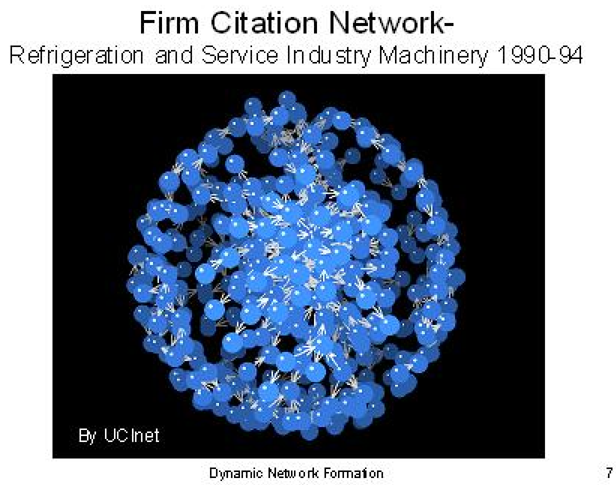


Figure 2.2: Mu-In and Log(Random Friends/Network-Based Friends)

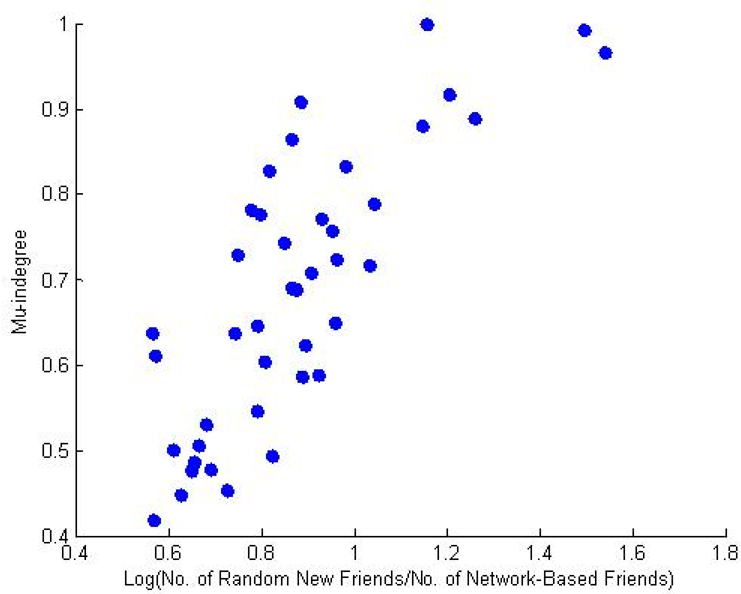


Figure 2.3: Mu-Out and Log(Random Friends/Network-Based Friends)

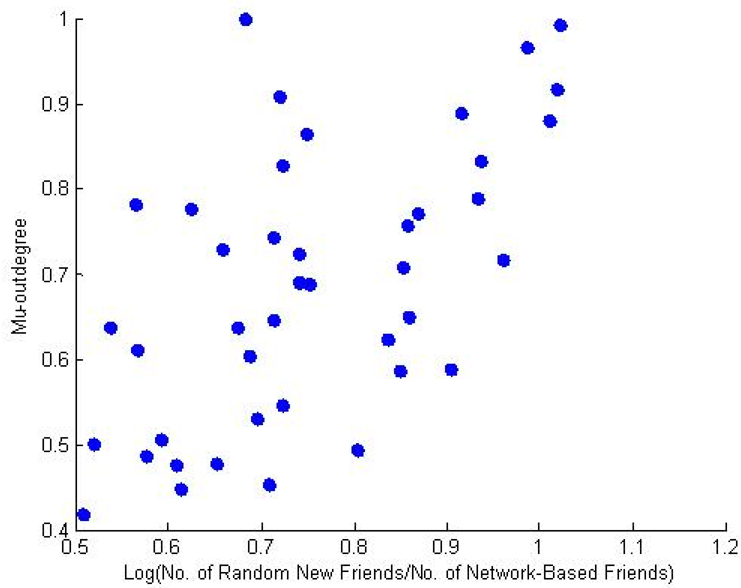


Figure 2.4: Mu-Total and Log(Random Friends/Network-Based Friends)

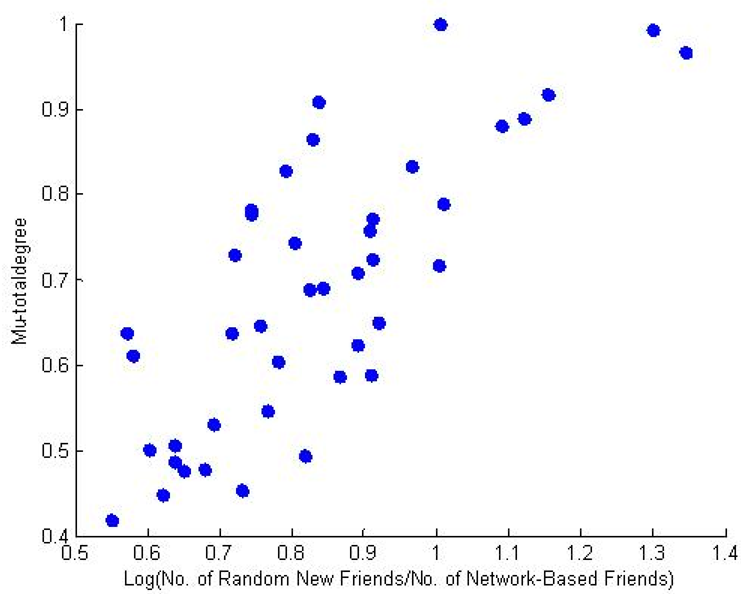


Figure 2.5: CC-TT and Log(Random Friends/Network-Based Friends)

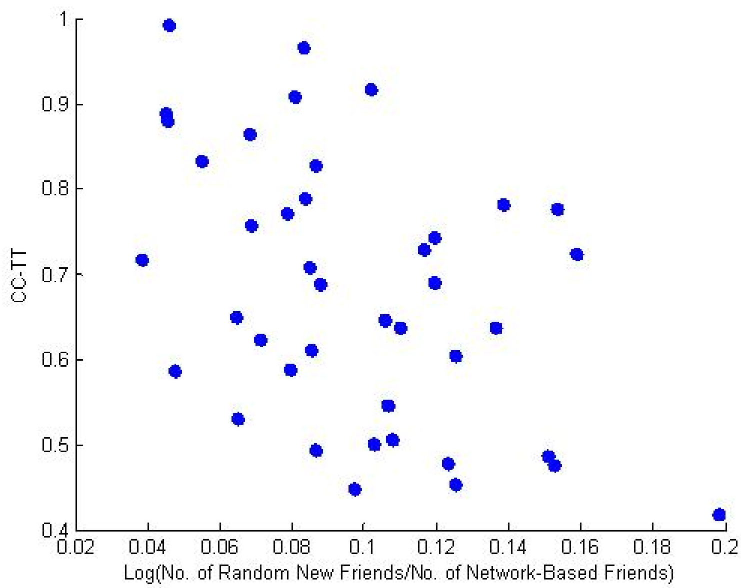


Figure 2.6: CC and Log(Random Friends/Network-Based Friends)

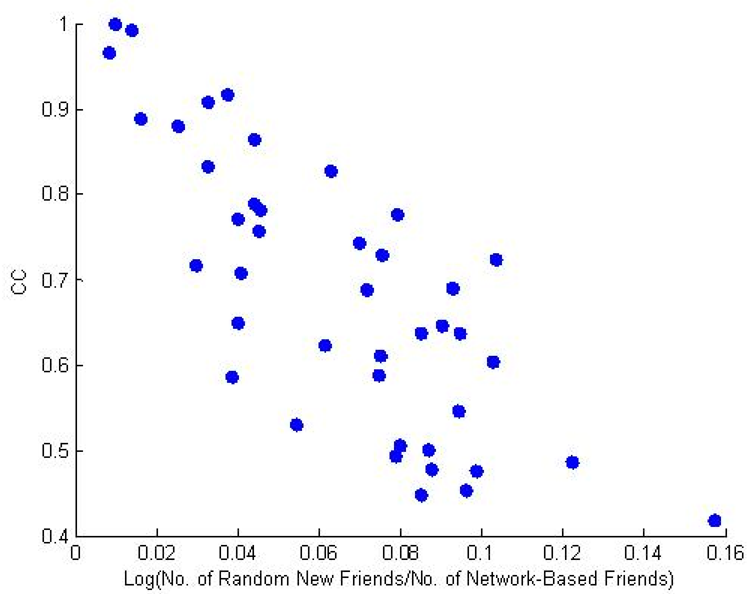


Figure 2.7: CC-Avg. and Log(Random Friends/Network-Based Friends)

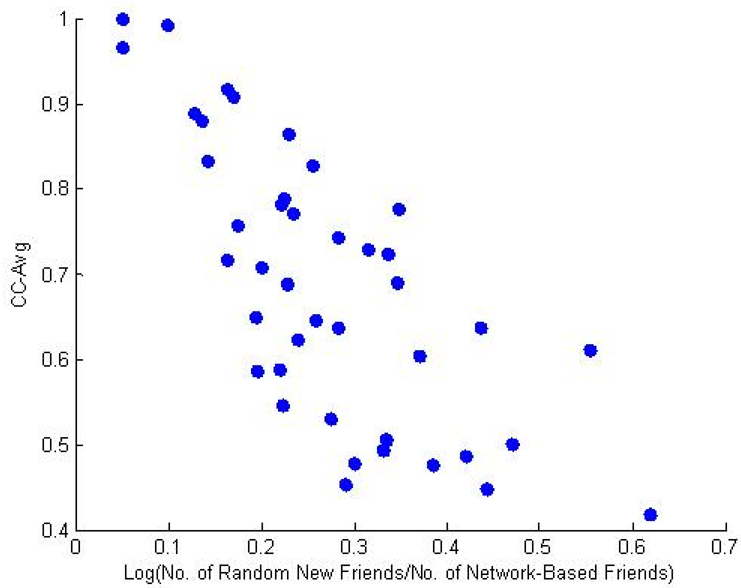


Figure 2.8: Mu-In of Real Networks and Mu-In of Simulated Networks

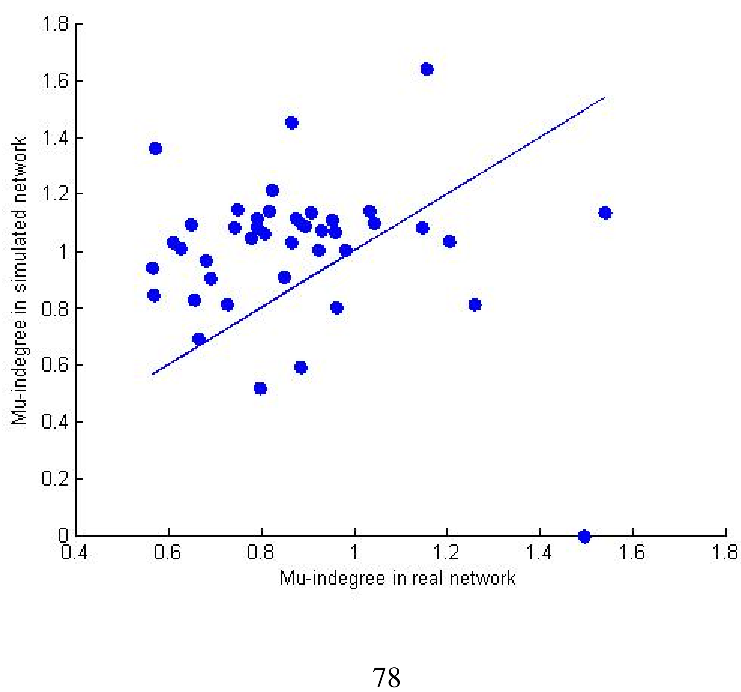


Figure 2.9: Mu-Out of Real Networks and Mu-Out of Simulated Networks)

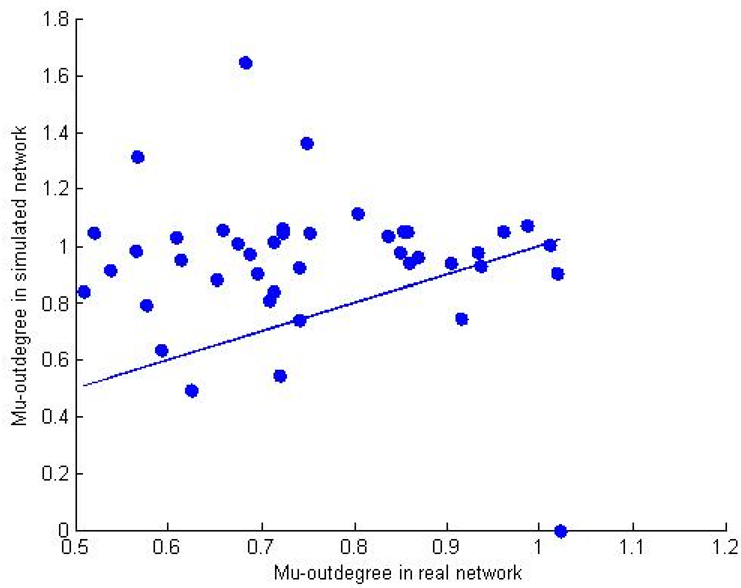


Figure 2.10: Mu-Total of Real Networks and Mu-Total of Simulated Networks

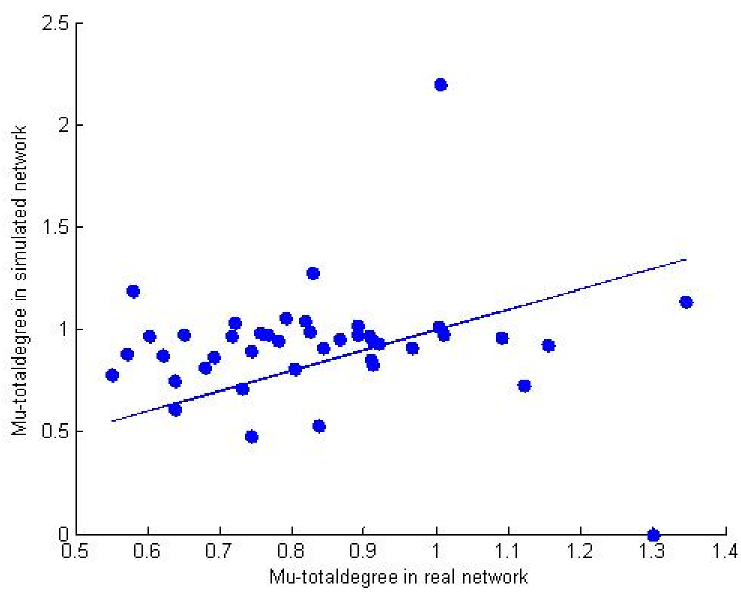


Figure 2.11: CC-TT of Real Networks and CC-TT of Simulated Networks)

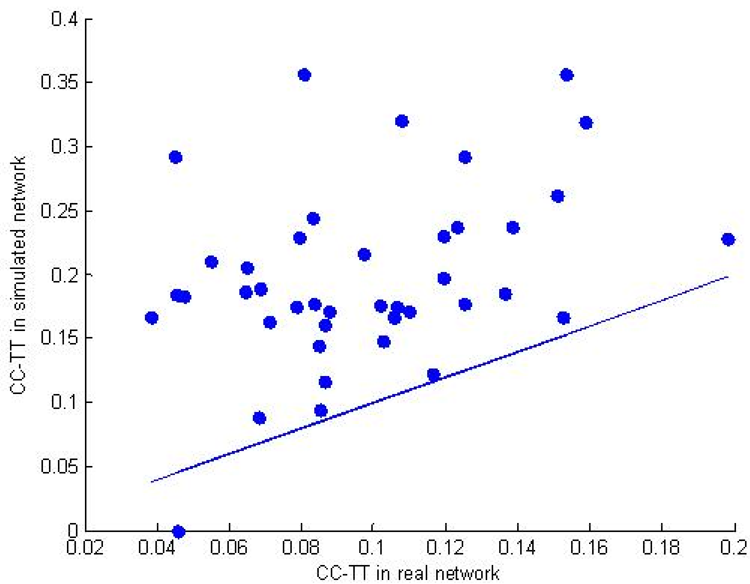


Figure 2.12: CC of Real Networks and CC of Simulated Networks

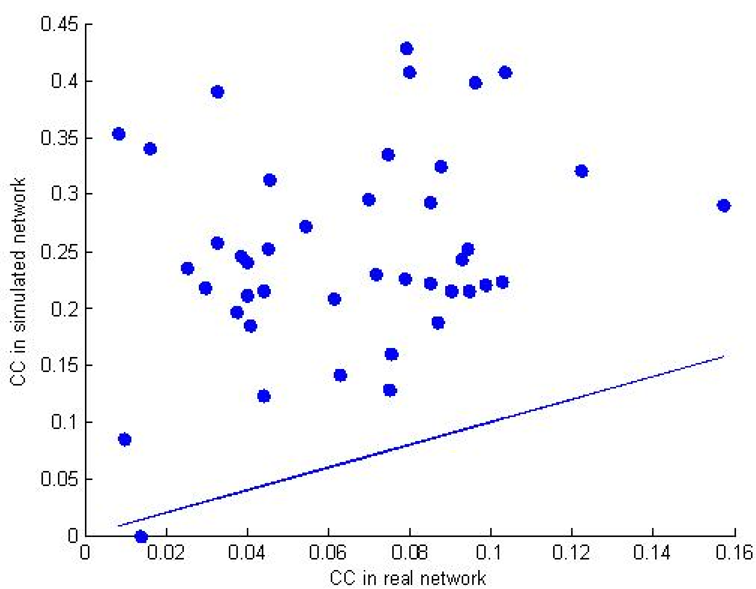


Figure 2.13: CC-Avg. of Real Networks and CC-Avg. of Simulated Networks)

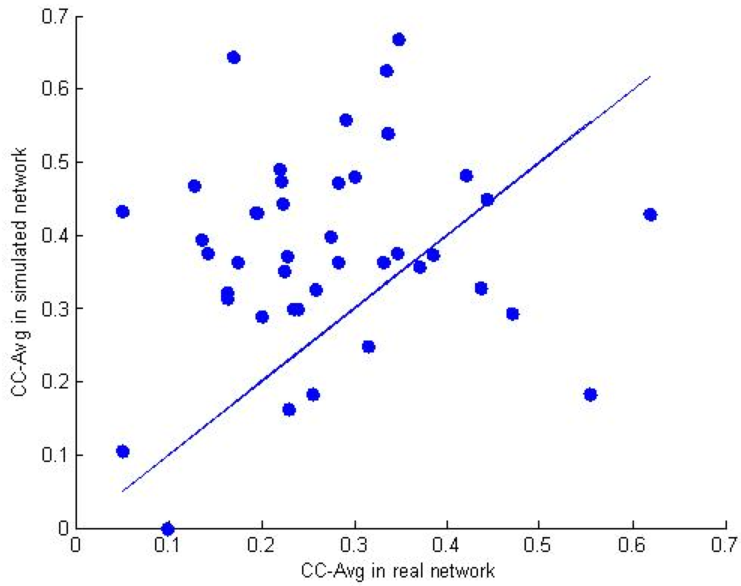


Figure 2.14: CC of Real Networks and CC of Simulated Networks

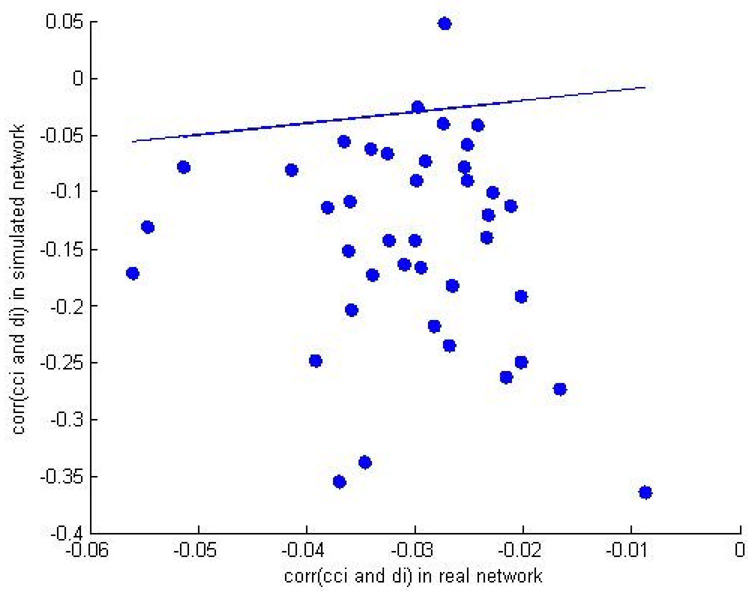


Figure 2.15: CC-TT and Mu-Out in Real Networks

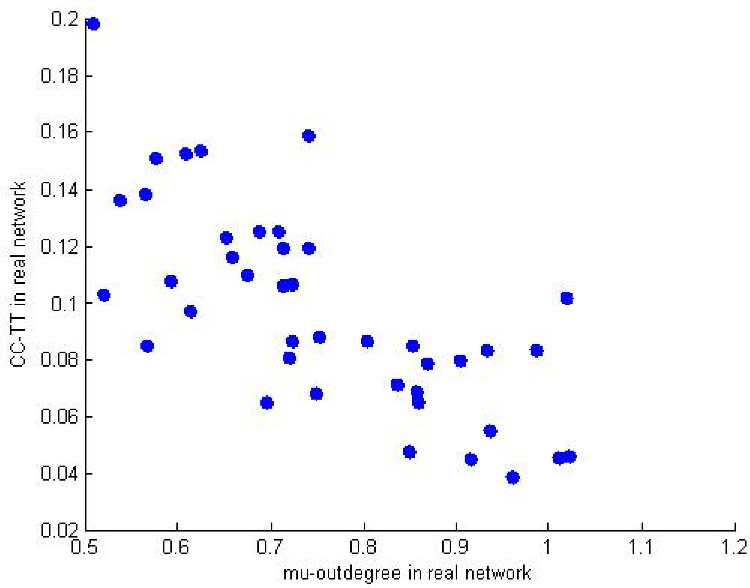
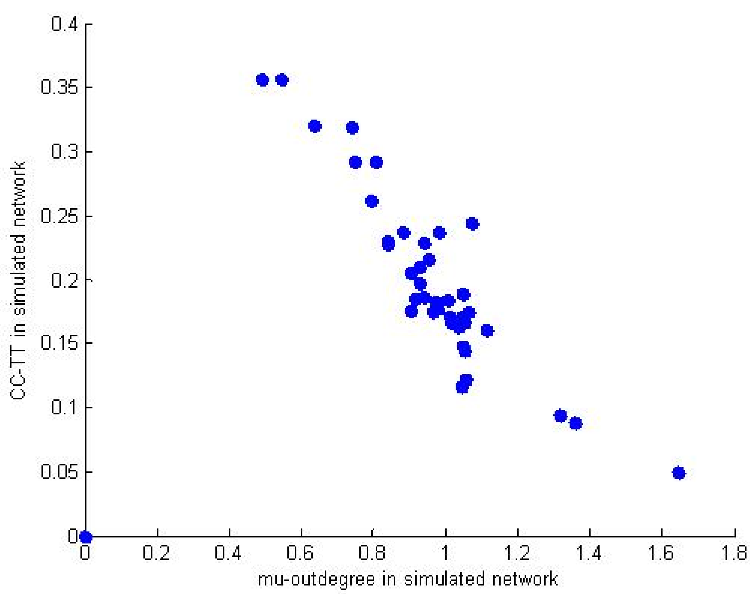


Figure 2.16: CC-TT and Mu-Out in Simulated Networks



Chapter 3

Information Heterogeneity by Firm Size and Aggregate Fluctuations

3.1 Introduction

This paper is motivated by two concurrent facts: first, information heterogeneity among firms increases after the 1980s; second, aggregate output becomes less volatile and more persistent after the 1980s. The term of each full business cycle also increases from 58.2 months during 1960-1980 to 109 months after the mid-1980s, as announced by the National Bureau of Economic Research (NBER) Official Business Cycle Dating Committee. When I measure the information updating speed by patent citation time lag in NBER Patent Data, I find that smaller firms cite older existing patents than larger firms. It is rational for smaller firms to update information more slowly than larger firms due to their limited process capacity. In a model with size-dependent reaction time lag and Pareto firm size distribution, the gradual cross-firm diffusion of a micro-level technology shock generates a persistent and hump-shaped aggregate output growth rate. Greater information heterogeneity across firms de-synchronizes the co-movement among firms of different sizes, and hence causes less volatile, more persistent and longer aggregate business cycles. Among the literature about information heterogeneity, the model with size-dependent reaction time lag to shocks is well suited to explain several facts about the business cycle's timing relation. For example, productivity dispersion is pro-cyclical, the top firms' growth rate predicts future GDP growth, investment leads hiring over the business cycle, and labor-share is counter-cyclical.

Patent citation data show that large firms acquire information faster than smaller firms due to their ability to access better and broader knowledge sources, obtain more updated information from the same knowledge source, and locate closer to R&D centers. From the 1980s to 1990s, information heterogeneity in terms of citation time lag widens in two ways: firms obtain knowledge from a more and more exclusive clique, and the citation time lag gap between large and small firms broadens. Meanwhile, the dynamic formation process of firm networks also changes; firms build new connections more often through the introduction of existing links, and are less likely to connect with a random unconnected firm. All these changes block knowledge from diffusing evenly across all firms and lead to greater information heterogeneity. Changes in intellectual property protection towards more strengthened protection may indirectly cause slower information diffusion across firms and greater information heterogeneity.

Combining the size-dependent reaction time lag with a Melitz (57) type model, I derive an analytic solution to illustrate information heterogeneity's impact on aggregate fluctuation. I assume a z-sized firm's reaction time lag follows Poisson distribution with average $\lambda(z) = c - d\log(z)$. Poisson lag distribution works well with *Pareto* firm size distribution, so there is an explicit solution. Without aggregate level Total Factor Productivity shock, I show that the gradual diffusion of a firm-level productivity shock[1] can generate a hump-shaped and persistent aggregate growth rate. *Pareto* firm size distribution is also important for generating an evenly hump-shaped impulse response. *Pareto* distribution's $p.d.f$ $f(x) = \mu x^{-\mu-1} x_{\min}^{\mu}$ [2] is special, because the product of a given firm size x and the density of the firms of x size $f(x)$ is constant over x for $\mu \approx 1$. During the periods after the shock, there

[1]Firm productivity shock comes from the shocks to firm level innovation and the production process. For example, a successful process innovation is a positive shock, while a strike is a negative shock.

[2]$x_{\min}$ is the minimum size of a firm. For example, the smallest firm contains one worker, the owner his/herself.

is always a similar share of total production affected by the shock. Other distributions, for example log-normal distribution, with smaller density at the lower end of firm size distribution, still generates a hump-shaped impulse response function, but the response function plunges shortly after the peak, instead of persistently and slowly decreasing to zero.

Compared to previous literature about information heterogeneity, the model featuring cross-firm differences in reaction time lag is well designed to explain several timing relations relevant to the business cycle. For example, Comin and Mulani (16), Comin and Gertler (15), and Davis, Haltiwanger, Jarmin, and Miranda (19) show that cross-section firm productivity dispersions are pro-cyclical over business cycles. When positive technology shock hits the economy, large firms upgrade their technology more quickly than small firms; the productivity gap between large and small firms widens temporarily until small firms also learn the improved technology at a later date. In contrast, when the economy is hit by a negative technology shock, large firms contract earlier than small firms, and therefore the productivity gap between them shrinks temporarily until the small firms are also influenced adversely. Gabaix (28) shows that the top 100 firms' growth rate predicts future GDP growth rate. In this model, large firms' growth rate leads the GDP simply because these firms react to shocks faster than the rest of the economy. For example, when a new technology emerges, large firms adopt it earlier than small firms. As a result, the aggregate growth rate reacts to technology shock with longer delays than in top firms.

When I vary the degree of information heterogeneity d, I show that greater information heterogeneity reduces aggregate volatility by desynchronizing co-movement among firms of different sizes. The term of one full business cycle also becomes longer. Since the cycle ends when the smallest firms finally adopt the new technology, it takes longer for the smallest firm to adopt the new technology when information diffuses more slowly (d is larger). In addition, top firms' growth rates lead the GDP for longer periods and pro-cyclical productivity dispersion becomes more significant in the simulations with larger information heterogeneity d.

In future work, I will improve this model in the following ways. First, firms in the model are pure competitors, without roles of complementariness through input-output matrix or specialization and cooperation. The simple competition relation limits the number of shock propagation channels in the model. Second, the citation time lag is measured in years, which is too long to study monthly or quarterly fluctuations. I have to assume that the relation between reaction time lag and firm size still holds at shorter time horizons. In addition, there is no capital in the model, and it therefore cannot capture capital-related stylized facts about the business cycle, such as the lead and lag relation between investment and hours, or the counter-cyclical labor share in total value added.

3.2 Literature

There are two streams of literature related to this topic. One stream concerns information 'stickiness' and its macroeconomic implications, for example, see Carroll (12), Mankiw and Reis (56), Sims (65), Woodford (70) and Moscarini (58), Ball, Mankiw, and Reis (6), Reis (61), Luo (52), and Abel, C., and Panageas (1). They point out that an individual is too small to gather and analyze all available information. Therefore, information stickiness exists, because agents choose to be inattentive for some interval. Carroll (12) also shows that agents from different demographic groups update information at different intervals and form heterogeneous expectations about the economic status. The other papers, however, assume that all agents are ex ante identical in information process capability and that the ex post information heterogeneity is completely due to the randomness of private signals. In contrast, information heterogeneity across firms in this paper originates from firms' various abilities to absorb and gain from updated knowledge, conditional on firm size.

The other literature stream uses idiosyncratic firm- or sector-level fluctuations to explain aggregate shocks; for example, see Jovanovic (42), Durlauf (24), Bak, Chen, Scheinkman, and Woodford (5), Nirei (60), Gabaix (27), Long and Plosser (50), Dupor (23), Horvath (36), Horvath (37) and Conley and Dupor (17). The structure of the input-output matrix (Long and Plosser (50),Horvath (36) and Dupor (23)), non-linear interaction between firms (Jovanovic (42)), non-convex cost technology (Bak, Chen, Scheinkman, and Woodford (5)), and the type of technology shock matters (Comin and Mulani (16)) when studying the impact of independent micro-level shocks on aggregate fluctuations. Gabaix (28) shows that

Pareto firm size distribution breaks the law of large number, when the macro shock comes from the aggregation of individual firm-level shocks. The granular nature of the economy determines that idiosyncratic shocks to a number of top firms are sufficient to explain the part of aggregate-level fluctuation.

In this model, the shock propagation channel is the firm's size-dependent reaction time lag. An exogenous firm-level technology shock diffuses gradually across firm networks. Large firms update information more often than smaller firms, because their processing cost is lower or their gain from newer knowledge is greater. *Pareto* firm size distribution helps generate a persistent and even hump-shaped impulse response. Furthermore, the special ability of this model is to explain the timing relations during the business cycle.

3.3 Information Heterogeneity In Patent Citation Data

Information heterogeneity is defined here by the speed of knowledge acquisition and processing. Specifically, in patent citation data the speed is represented by the citation time lag, which is equal to the application year of the citing patent minus the application year of the cited patent. Smaller firms have a disadvantage in acquiring and absorbing information compared to larger firms, in that they cite older patents than those cited by larger firms.

The information heterogeneity observed in the NBER Patent Citation Data provides empirical evidence relevant to the literature on information heterogeneity and rational inattention. Slower learning can be a smaller firm's optimal choice, because the cost of obtaining the most up-to-date information outweighs the benefit, given the limited capacity of small firms to process information. Smaller firms may also benefit less from fresh news than large firms do. Since information processing cost is more or less a fixed cost, large firms can cover the fixed cost by applying it to a wider scope of products.

In detail, information heterogeneity (IH) among U.S. firms encompasses three aspects: the chances to learn from better sources, the speed at which it is possible to learn from the same source, and the geographic distance to research centers. First, firms choose to learn from peers similar to themselves. Firms make more citations to peers in their own sector[3]. Larger firms tend to cite larger firms[4]. Better connected firms tend to cite better connected firms[5].

These facts echo the assortative mixing exhibited among other types of social networks documented in Newman (59). This pattern arises if the cost of absorbing external knowledge increases with the gap between the learner and the source. As a result, on average, larger firms learn from better knowledge sources than smaller firms. Second, among the citing firms that cite patents owned by the same cited firm, larger citing firms tend to cite newer patents than smaller citing firms. This supports the idea that larger firms are faster than smaller firms in acquiring and adopting outside knowledge, given that they are able to reach the same knowledge source. Third, firms located more closely together learn from each other more quickly. Firms within the same city and the same state cite each other faster than they cite other firms. Unfortunately, smaller firms are less likely than large firms to locate in the top R&D centers. In summary, larger firms have an advantage over smaller firms, in that they can reach larger knowledge sources, acquire more current knowledge from the same source, and locate closer to research centers.

The information heterogeneity examined among firms in this paper differs from the information stickiness in previous literature. Here, different agents update information at different frequencies, according to their processing capacity. This is similar to the Carroll (12) model, where households from different demographic groups form heterogeneous inflation expectations due to their different frequencies of updating information.

3.3.1 Greater Information Heterogeneity after 1980s

The information heterogeneity among U.S. firms became stronger during the 1980s. The citation lag differences due to firm size, distance and sector border enlarge over time, due to the change of network structure[6]. Since citation lag is a nonnegative

[3]67% of citations made between U.S. organizations are within the same sector.

[4]Correlation between the log scale patent numbers of the citing and the cited firm is 0.25, when both firms belong to the same sector.

[5]Correlation between the log scale hub weights of the citing and the cited firm is 0.38; correlation between the log scale authority weights of the citing and the cited firm is 0.21.

[6]Since one firm may cite a patent multiple times, I include only the citations that represent when the citing firm cites a given patent for the first time. The first-time citation indicates the earliest time

integer, which counts the number of continuous periods without successful search of information, I use a Cox Proportional Hazards model to estimate the likelihood of a successful search per period. Observations are grouped by patent class.

The specification of the Cox regression is as follows

$$log(HR_t) = b_{0,t} + b_{1,t} \ln(patent\_no_{citing,t}) + b_{2,t} samecity + b_{3,t} samestate + b_{4,t} cross + c\_patent\_class$$

HR_t is the hazard rate of a successful search per period. $\ln(patent\_no_{citing,t})$ and $\ln(patent\_no_{cited,t})$ are the number of patents owned by the citing firm and cited firm at year t, respectively. $ln(dist)$ is the log-scale great circle distance in kilometers between the inventors of the citing patent and the cited patent. $samecity$ and $samestate$ are dummy variables indicating whether the citing firm and the cited firm are located in the same city and same state. $cross$ is a dummy variable indicating whether the citing patent and the cited patent belong to different sectors. c_patent_class is the sector or patent class fixed effects. The citations are grouped by the citing year. There is one such regression per year. Among the firms that applied patents in the same patent class, hazard rate is larger (citation time lag is shorter), if the citing firm is larger, the citing firm is located closer to the cited firm, and the two patents belong to the same patent class. The regression results are reported in Table 3.1.

More importantly, the hazard rate or citation lag differences due to size and location grew over time during the 1980s and 1990s. Suppose firm A is 10 times larger than firm B, they are located in the same city and apply patent in the same industry: firm A is 4.04% more likely to find the right knowledge per period than firm B in 1985, while in 2000, firm A's hazard rate of a successful search is 12.14% higher than that of firm B's. Despite the improving telecommunication technology, the obstacle of distance became stronger. In 1985, locating in the same city and the same state as the cited firm increased the hazard rate by 3.0% and 2.1%, respectively, while in 2000, the differences enlarged to 14.9% and 13.8%. The hazard rate difference due to sector border increased slightly, but is stable around 10%. Large

that the citing firm learns from the cited patent. First-time citations also mark new knowledge flow between firms.

firms learn much faster than smaller firms. Knowledge takes longer to spread to faraway places.

From another perspective, the dynamic formation process of firm networks during the same period also demonstrates that information is shared less evenly across firms. Each firm shares more information with closely connected peers and less with random, disconnected firms. In detail, in every period one firm may cite other firms that it did not cite the previous year. I divide these new connections into two types following Jackson and Rogers (38): friend's friend (FF) and random friend (RF). The ratio between FF/RF rises steadily from 0.6 in 1984 to 0.77 in 1995. That is, when searching for a new knowledge source, a firm relies more and more on existing connections; the chance to randomly build a link with an unknown firm is declining[7]. This trend means that firms are more constrained when looking for new knowledge sources. As a result, information is shared within a small, already well connected neighborhood, and communication with random outsiders declines. Information heterogeneity across the entire network becomes greater over time.

Overall, all citation lags became longer relative to the age of patent stocks after the 1980s. For instance in Figure 3.1, if I use the ratio between citation share and patent stock share to measure the inward citation intensity for patents of different ages, in the 1980s firms cited the most recent patents over-proportionally, whereas in the 1990s firms cited the most recent patents under-proportionally[8].

3.4 The Model

This model is based on an autarky version of Melitz (57).

3.4.1 Consumer

The representative consumer faces the following problem:

[7] The cause of this phenomenon could be that firms become more specialized, so they only find information from closely related firms useful. It could also happen if the communication technology has changed in such a way that it strengthens the tie between already connected firms, but doesn't help to build trust between two unknown firms.

[8] The citations used in this figure include only first-time cross-firm citations.

$$U = \max_{\{x_{i,t}\}} \int_0^\infty \rho^t \left[\log\left(Y_t\right)\right] dt \tag{3.1}$$

$$Y_t = \left(\int_0^1 x_{i,t}^{\frac{\sigma-1}{\sigma}} di\right)^{\frac{\sigma}{\sigma-1}}.$$

ρ is the time preference of the representative consumer; Y_t is the consumption of final goods; x_{it} is the consumption of intermediate good i; product i is sold at price $p_{i,t}$. P_t is the aggregate price index. The mass of intermediate goods firms is 1. $\sigma > 1$ is the elasticity of substitution between intermediate goods. Consumer demand for intermediate goods is

$$x_{i,t} = Y_t \left(\frac{P_t}{p_{i,t}}\right)^\sigma$$

$$P_t = \left(\int_0^{I_t} p_{i,t}^{1-\sigma} di\right)^{\frac{1}{1-\sigma}}$$

3.4.2 Heterogeneous Firms

In a monopolistic competitive market, there are numerous intermediate goods firms with various sizes due to heterogeneous labor productivity. The mass of firms is 1. Productivity distribution follows *Pareto* distribution $F\left(A_{i,t}\right) = 1 - \left(\frac{A_{i,t}}{A_{\min}}\right)^{-\mu}$. $A_{\min}$ is the smallest productivity able to survive due to fixed operation cost. μ is the shape parameter, with larger μ meaning firm size distribution is more homogeneous. A firm's productivity is constant overtime unless it is hit by technology shocks. Firms can flexibly adjust their prices according to their productivity at any time.

Production uses only labor. Firm i's production function is

$$x_{i,t} = A_{i,t} L_{i,t}.$$

Wage rate is normalized to 1. Firm i's optimal price is

$$p_{i,t} = \frac{\sigma}{(\sigma-1) A_{i,t}}.$$

The aggregate price level is

$$P_t = \tilde{p}_t = \frac{\sigma}{(\sigma-1)\tilde{A}_t},$$

which is equal to the optimal price of a firm with productivity $\tilde{A}_t = \left(\int_{A_m}^{\infty} A_{i,t}^{\sigma-1} dF\left(A_{i,t}\right)\right)^{\frac{1}{\sigma-1}} = \left(\frac{\mu}{\mu-\sigma+1}\right)^{\frac{1}{\sigma-1}} A_m$, $\mu_A - \sigma + 1 > 0$ to ensure that the $\tilde{A}_t$ is finite. $\tilde{A}_t$ represents the quantity weighted average firm productivity.

Firm i's production becomes

$$x_{i,t} = Y_t \left(\frac{A_{i,t}}{\tilde{A}_t}\right)^{\sigma}.$$

Firm size heterogeneity originates from productivity heterogeneity.

In addition to productivity dispersion, firms are also heterogeneous in their network connectivity. Cross-firm patent citations link all firms as a network. The number of other firms with which a firm is directly connected is called the degree of that firm. The degree distribution across firms follows a Power-law distribution, which is the discrete version of *Pareto* distribution. A firm that owns more patents also has more linkages in the network.

Size Dependent Reaction Time Lag

Information dispersion associated with firm size heterogeneity is embodied in the model by the declining reaction time lag with firm size. Specifically, the time lag that firm f with productivity $A_{f,t}$ requires to react to a news is a random draw from a Poisson distribution with parameter $\lambda_t\left(A_{f,t}\right) = c_t - d_t \ln\left(A_{f,t}\right)$. Larger d_t means larger firms react to news much faster than smaller firms, hence greater information heterogeneity in the economy. $c_t = c_0 + d_t \ln\left(A_{\max}\right)$[9] to ensure that even the largest firm with productivity $A_{\max}$ takes c_0 time to gather information. If $A_{\min} = 1$, the smallest firm's expected time to hear the news after c_t periods.

[9]According to Newman (59), length (in number of edges) of the longest geodesic path between any two nodes in a network increases as fast as $\ln(N)$, where N is the number of nodes in the network. When the largest firm's productivity $A_{\max}$ is proportional to the number of firms N or population size, the assumption that $c_t \propto \ln(A_{\max})$ also means the longest reaction time in the network $\lambda\left(A_{\min}\right) = c_t \propto \ln(N)$.

If every firm processes information at the same speed $d_t = 0$, c_t is the reciprocal of information updating frequency, or $\frac{1}{\mu}$ in Reis (61). The Poisson distribution of reaction time lag works well with *Pareto* distribution of firm size, in the sense that there is an analytical solution for total production k periods after the shock.

The origin of a positive technology shock is an individual firm's innovation and imitation activity. An example is a successful new innovation of a general purpose technology. The inventor of this new technology is the first firm to apply this new technology. All the firms that are directly linked to the inventor firm apply it in the first wave; those that are indirectly connected to the inventor firms apply it with longer lags, depending on their distances from the inventor. If one measures each patent's impact by the number of citations it receives, Hall, Jaffe, and Trajtenberg (31) show that the distribution of each technology shock's impact is as skewed as a *Pareto* distribution. That means only very few new technologies are able to attract attention from all peers and eventually induce macro-level fluctuations.

A negative productivity shock could happen when some new distraction decreases employee's productivity at work; for example, instant messaging software and social network web sites. It also could happen because of temporary infrastructure dysfunction, a strike, or an infectious disease. A larger firm is also more likely to receive a negative productivity shock earlier than a small firm, because it hires more workers, each of whom is equally like to introduce a new distraction or infectious disease to the company; a large firm also generates more tasks that demand public service and high quality infrastructure. Gabaix (28) also uses strike or the threat of a potential strike as a firm-specific negative productivity shock.

Suppose a 1% positive technology shock hits the economy at time t. Denote the share of firms with productivity A that are just-informed at time $t+k$ as

$$JI(k,A) \equiv \frac{\lambda_t(A)^k e^{-\lambda_t(A)}}{k!} = \frac{[c_t - d_t \ln(A)]^k A^d}{k! e^{c_t}}.$$

By definition, $JI(k,A)$ is the probability mass function of Poisson distribution with

$\lambda(A) = c_t - d_t \ln(A)$. By $t+k$, the share of just informed firms in the economy is

$$\begin{aligned} JI(k) &= \int_{A_{\min}}^{\infty} JI(k,A)\, dF(A) \\ &= \int_{A_{\min}}^{\infty} \frac{(c_t - d_t \ln(A))^k A_m^{\mu} \mu A^{d_t - \mu - 1}}{k! e^{c_t}} dA \\ &= \frac{\mu}{\mu - d_t} \sum_{g=0}^{k} \left(-\frac{d_t}{\mu - d_t}\right)^{k-g} JI(g, A_{\min}). \end{aligned}$$

At time $t+k$, the share of all informed firms is:

$$I(k) = \sum_{g=1}^{k} JI(g)$$

.

$JI(k,A)$ and $I(k)$ summarize the aggregate status of information diffusion among firms. When $d_t = 0$, $JI(k) = JI(k, A_{\min})$, which is the probability mass function of Poisson distributions with $\lambda = c_t$. Each firm has the same chance to learn the news as the smallest firm at time $t+k$. $JI(k, A_m)$ is the probability density function (p.d.f.) of Poisson distribution with parameter $\lambda = c_t - d_t \ln(A_{\min})$. For example, when $A_m = 1$, $d_t = 0$. Very slow information transmission within the firm network (large λ) means $JI(k, A_{\min})$ and $I(k)$ is smooth and symmetric over k, because every period only a tiny group of the firms is informed.

3.4.3 Aggregate Fluctuation under Information Heterogeneity

In this subsection, I vary the degree of information heterogeneity d to illustrate its impact on several aspects of aggregate fluctuation. First, larger d causes a more persistent, less volatile and longer hump-shaped impulse response of total output. Second, larger d induces a situation in which the top 100 firms' growth rate leads GDP growth rate for more periods. Third, stronger information heterogeneity incurs greater pro-cyclical cross-sectional firm productivity dispersion.

Hump-Shaped Impulse Response

Firms adjust their labor input and productivity according to the information they possess; therefore the shape of $JI(k,A)$ and $I(k)$ govern the cyclical pattern of aggregate production and productivity dispersion over a cycle. Suppose the news is a positive technology shock: informed firms can produce at $1+\triangle$ times their original productivity, while uninformed firms still produce at their original productivity. Firms informed early enjoy excess profit, until all firms learn this news and every firm's profit returns to the normal level justified by its initial relative productivity. The benefit of learning of a positive productivity shock earlier is twofold: a firm enjoys excess profit for a longer duration, and the excess profit is much higher in earlier periods, because the average productivity is still low because few firms are aware of the better technology.

The aggregate production at t+k depends on how many firms are informed and how large these firms are

$$\begin{aligned} Y_{t+k} &= \frac{1}{P_{t+k}} = \frac{(\sigma-1)\tilde{A}_{t+k}}{\sigma} \\ &= \frac{(\sigma-1)}{\sigma}\left\{ \begin{array}{c} \frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g} JI\left(g,A_{\min}\right) \\ +\frac{\mu}{\mu-\sigma+1} \end{array} \right\}^{\frac{1}{\sigma-1}} A_m \end{aligned}$$

Y_{t+k} and the growth rate of Y_{t+k}, g_{t+k}, have similar curvature as the mass of informed firms $I(k)$ and the mass of just-informed firms $JI(k)$, respectively. See Figure 3.2. With information heterogeneity or $d_t > 0$, g_{t+k}, exhibits a hump-shaped curve over k. The growth rate at the beginning comes mainly from the just-informed top firms; each of them is a giant but their mass is small in *Pareto* distribution. At a later stage, $JI(k)$ rises first and drops later, but the average size of just-informed firms always declines. The peak of g_{t+k} comes when the product of just-informed firms $JI(k)$ and their average size reach the maximum. In contrast, when $d_t = 0$ there is no hump-shaped impulse response of growth rate, because when every firm faces the same time lag distribution, the average size of just-informed firms are constants over time, and the growth rate plunges simply as $JI(k)$ declines over time.

Here the slow propagation of positive technology shock is completely due to slow transmission of information across firms. Greater information heterogeneity (larger d_t) de-synchronizes the co-movement between firms of different sizes. As a result, the peak of the $\{g_{t+k}\}$ curve comes later and lower, when d_t is larger.

The variance of $\{g_{t+k}\}$ over $k = 0$ to 30 also decreases with information heterogeneity d_t, even though the size of the technology shock is the same. This model may provide an additional explanation for the more than halved aggregate volatility or great moderation since the 1980s. Policy changes towards stronger intellectual property protection since the 1980s have induced slower knowledge diffusion across firms and the formation of highly clustered firm networks, which cause greater information heterogeneity among firms. In consequence, the macroeconomic fluctuations become less volatile when firms move at de-synchronized paces.

Lead and Lap between Top Firms' Growth and GDP Growth

Gabaix (27) shows that the growth rate of the top 100 non-oil and non-energy public U.S. firms predicts GDP growth rate. He lists several potential explanations: autocorrelation at the firm level; imitation dynamics, where a successful technology is imitated by other firms; time aggregation; and the propagation of shocks along supply and demand chains, as in Long and Plosser (50). This model proposes another explanation: the propagation of shocks through firm networks. Top firms learn of the shock earlier than other firms, and therefore their reaction leads the rest of the economy. This is similar to imitation dynamics and the propagation of shocks along supply and demand chains, which both emphasize the difference in reaction time across firms.

Analytically, the output of top firms whose productivity is higher than A_x, $A_x \ggg A_{\min}$, can be expressed as follows:

$$Y_{t+k}^{A>A_x} = \left(\frac{\tilde{A}_{t+k}|A_{t+k}>A_x}{\tilde{A}_{t+k}}\right)^{\sigma} Y_{t+k} = \frac{(\sigma-1)}{\sigma}\frac{\left[\int_{A_x}^{\infty} A_{i,t+k}^{\sigma-1} dF\left(A_{i,t}\right)\right]^{\frac{\sigma}{\sigma-1}}}{\int_{A_x}^{\infty} A_{i,t+k}^{\sigma-1} dF\left(A_{i,t}\right)}$$

$$= \frac{\left[\frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g} JI\left(g,A_x\right)+\frac{\mu}{\mu-\sigma+1}\right]^{\frac{\sigma}{\sigma-1}}}{\frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g} JI\left(g,A_{\min}\right)+\frac{\mu}{\mu-\sigma+1}} \frac{(\sigma-1)A_x^{\sigma}}{\sigma A_{\min}^{\sigma-1}}$$

Top firms' growth leads GDP growth, intuitively because top firms react to productivity shock faster than other firms, and mathematically because the peak of $JI\left(g,A_x\right)$ leads that of $JI\left(g,A_{\min}\right)$. In Figure 3.3, I set $ln\left(A_x\right) = 18$, $ln\left(A_m\right) = 0$, $\mu = 1.06$[10], $\sigma = 1.05$, $\triangle = 1\%$ and $d = 0$, 0.1, 0.2, 0.4 in four scenarios. At the first few positive lags, the cross-correlation between $g_{t+k}^{A>A_x}$ and g_{t+k} is even negative for large enough information heterogeneity or d_t, because top firms squeeze other firms' market share when they update technology so much earlier than others. Not only does the positive cross-correlation come later for larger d_t, but the peak cross-correlation value is also lower. In contrast, when $d_t = 0$, top firms' growth and the growth of the rest of the economy completely coincide. Cross-correlation between them is simply self-cross-correlation of GDP growth rate.

Pro-cyclical Productivity Dispersion

Comin and Mulani (16), Comin and Gertler (15), Davis, Haltiwanger, Jarmin, and Miranda (19) show that cross-sectional firm productivity dispersions are pro-cyclical over business cycles. See Figure 3.4 and Figure 3.5 .[11]

In this model, pro-cyclical productivity dispersion is related to large firms' advantage in updating information. Since large firms are more likely to improve their productivity earlier than smaller firms, the productivity gap between large and

[10] Axtell (4).

[11] Figures 7 and 8 of Davis, Haltiwanger, Jarmin, and Miranda (19).

small firms widens temporarily until smaller firms learn the news (at a later date). The new cumulative density function (c.d.f) for productivity distribution becomes:

$$\begin{aligned}
\Pr\left(A_{f,t+k} > x\right) &= \Pr\left(A_{f,t} > x\right) + \int_{x/(1+\triangle x)}^{x} I(k,A)\, dF(A) \\
&\approx \left(\frac{x}{A_m}\right)^{-\mu} \left[1 + x\frac{\triangle x}{1+\triangle x} I(k,x)\right]
\end{aligned}$$

It is difficult to derive an analytic solution for the commonly used productivity heterogeneity measure, standard deviation of $\ln\left(A_{f,t+k}\right)$. Instead, I illustrate the productivity dispersion in a simulation with 100,000 firms. In Figure 3.6, with information heterogeneity due to firm size or $d_t > 0$, productivity dispersion's impulse response is also hump-shaped. The summit of the hump-shape comes later as d_t rises, because the productivity gap between early-informed large firms and later-informed small firms lingers longer. The summit is also higher as d_t rises, because the early-informed firms are more often large firms, and therefore the productivity distribution is skewed further right. In contrast, when $d_t = 0$, firm productivity dispersion is flat over time, because every firm has an equal chance to update productivity at a given time; the shape of the productivity distribution is well maintained, as it was before the shock.

Note that the productivity dispersion and GDP growth rate have similar pattern and timing for a given d_t. This echoes the empirical evidence that cross-sectional firm-level productivity dispersion and aggregate growth co-move at almost the same pace.

3.5 Conclusion and Future Research

This paper provides a potential explanation of the great moderation of aggregate volatility. Greater information heterogeneity across firms de-synchronizes the comovement across firms of different sizes, and hence reduces aggregate output volatility. The model with size-dependent reaction time lag is well suited to explaining the timing relations related to the business cycle; for example, the procyclical productivity dispersion and the lead-lag relation between top firms' growth rates and GDP growth rate.

Using patent citation data, I provide empirical evidence for information heterogeneity across firms measured by patent citation time lag. Larger firms are faster to obtain and process knowledge than smaller firms. In detail, larger firms are able to cite patents owned by larger and better connected firms; among the citing firms that cite patents owned by the same cited firm, larger firms cite more recent patents than smaller firms; and larger firms are more likely to locate closer to R&D centers. These facts support the literature of rational inattention, in that smaller firms update information less frequently, perhaps simply because of their poor processing capability or insufficient benefit from newer knowledge.

I analytically solve a simple Melitz (57) type model with *Pareto* distribution of firm size and size-dependent reaction time lag to shocks. The time lag with which a firm reacts to a technology shock follows a Poisson distribution, the mean of which is negatively related to the log-scale firm size. The model generates persistent hump-shaped aggregate fluctuation without aggregate TFP shocks. In the simulations with various degrees of information heterogeneity, greater information heterogeneity causes a less volatile GDP growth rate, a longer leading period of top firms' growth rate over GDP, and more pronounced pro-cyclical productivity dispersion.

In the future, I would like to incorporate more aspects of firm heterogeneity and explore the model's ability to match other stylized facts related to the business cycle. For example, larger firms use capital more intensively than smaller firms. The capital share difference can be as large as 25% within a three-digit level industry[12].

With this additional feature, this model will be able to illustrate why investment leads hiring over business cycles and why labor share is counter-cyclical. To expand production by the same amount, large firms invest more capital, while small firms hire more hours of labor. When large, capital-intensive firms move before small, labor-intensive firms, the aggregate investment leads hours over the business cycle. Meanwhile, when large, capital-intensive firms undertake a larger share of output during the boom and produce a smaller share during the recession, then at the aggregate level, labor receives a smaller share of value added during the boom and a larger share during the recession. The extension highlights that the

[12] See Young (71), Ambler and Cardia (3), and Hansen and Prescott (32).

model is suitable for capturing more timing relations over the business cycle.

Table 3.1: Cox Proportional Hazards model 1985-2000; Group Variable: Patent Class

Dependent Variable: Hazard Rate of Successful Search				
Year	1985	1990	1995	2000
Log(Patent Stock$_{Citing}$)[a]	.011***	.018***	.028***	.033***
	(.002)	(.002)	(.001)	(.001)
Samecity[b]	.030*	.054***	.109***	.149
	(.016)	(.011)	(.007)	(.005)
Samestate[c]	.021	.072***	.124***	.139
	(.029)	(.022)	(.014)	(.010)
Cross-sector citation[d]	-.092***	-.079***	-.084***	-.100
	(.009)	(.007)	(.004)	(.003)
No. of Observations	53272	98024	212902	368603
Wald chi2(4)	139.51	347.83	1684.72	4683.93

[a] Number of patents owned by the citing firm

[b] 1 if the citing inventor and the cited inventor live in the same city, 0 otherwise.

[c] 1 if the citing inventor and the cited inventor live in the same state, 0 otherwise.

[d] 1 if the citing patent and the cited patent belong to the same sector, 0 otherwise.

Figure 3.1: The Changing Ratio of Citation Share to Patent Stock Share

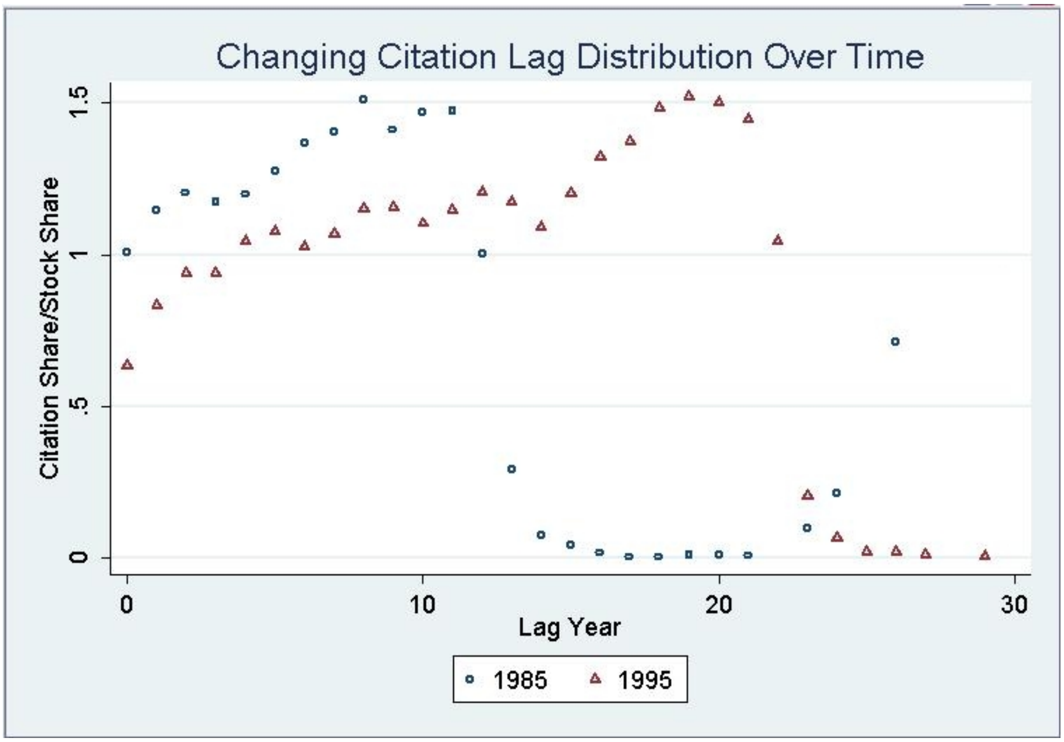


Figure 3.2: Information Heterogeneity and Growth Rate IR to Technology Shock

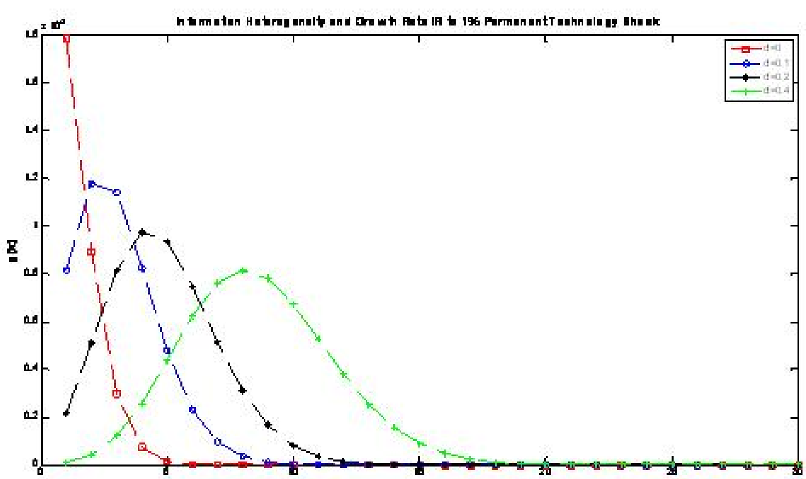


Figure 3.3: Figure 7 in Davis, Haltiwanger, Jarmin, and Miranda (19)

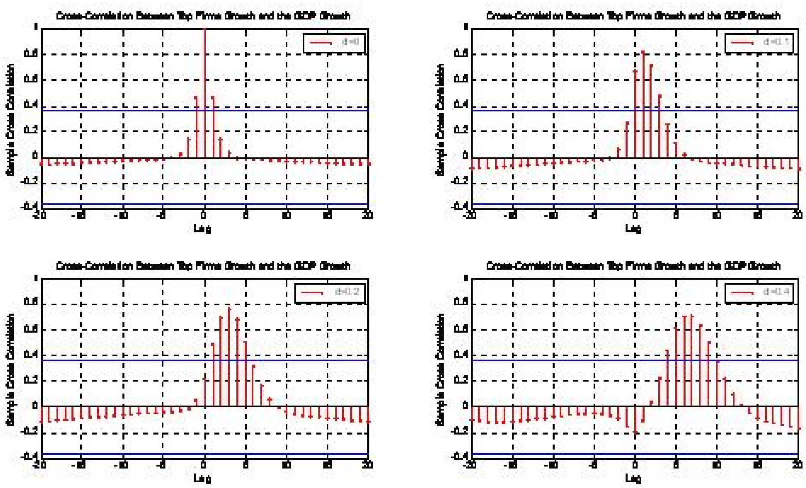


Figure 3.4: Cross-Correlation between Top Firms' Growth and GDP Growth

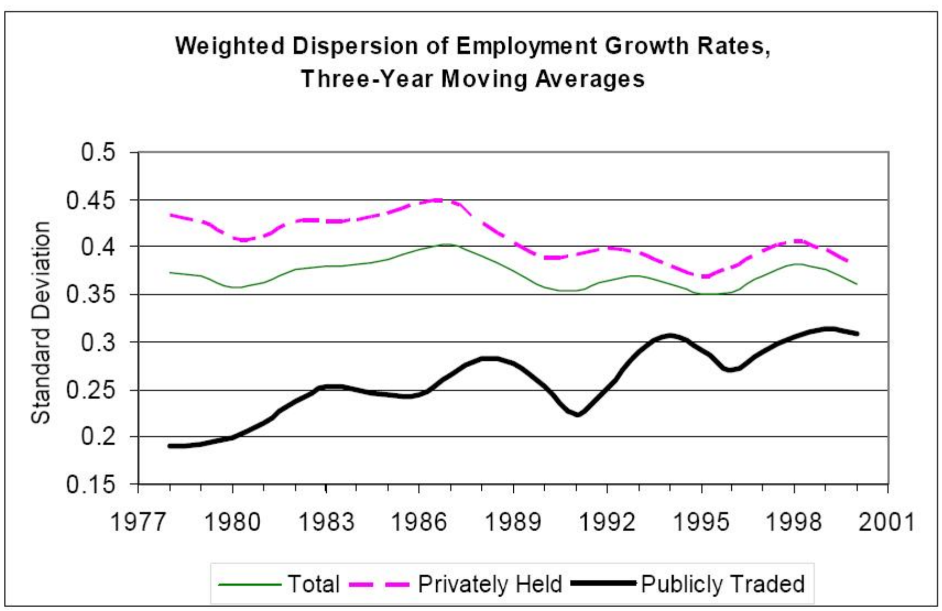


Figure 3.5: Figure 8 in Davis, Haltiwanger, Jarmin, and Miranda (19)

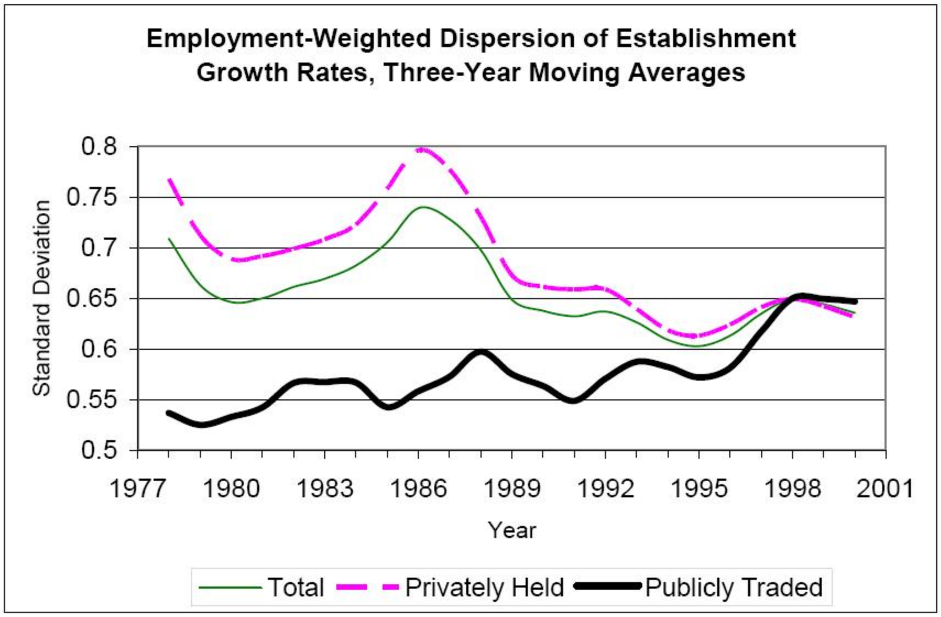
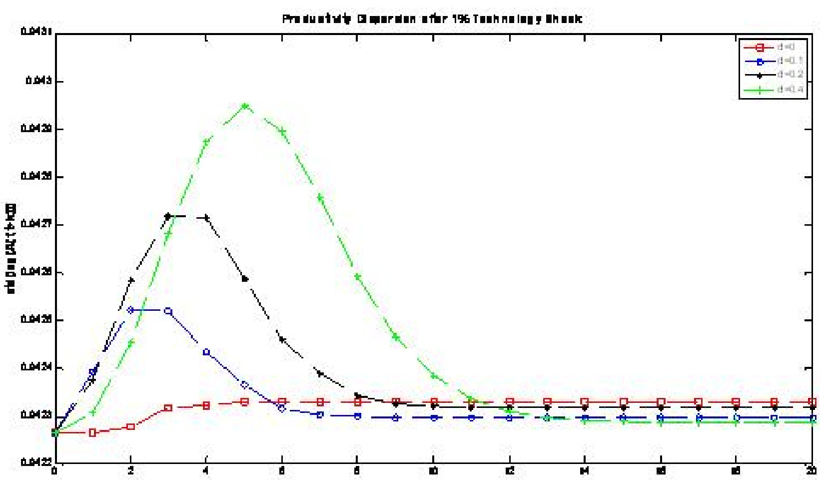


Figure 3.6: Productivity Dispersion after Technology Shock



Bibliography

[1] A. B. Abel, J. E. C., and S. Panageas. Optimal inattention to the stock market with information costs and transactions costs. NBER Working Paper 15010, National Bureau of Economic Research, Inc, 2009. → pages 86

[2] R. Albuquerque and H. A. Hopenhayn. Optimal lending contracts and firm dynamics. *Review of Economic Studies*, 71:285–315, 2004. → pages 7

[3] S. Ambler and E. Cardia. The cyclical behaviour of wages and profits under imperfect competition. *Canadian Journal of Economics*, 31(1):148–164, February 1998. → pages 99

[4] R. L. Axtell. Zipf distribution of u.s. firm sizes. *Science*, 293:1818–1820, 2001. → pages 6, 97

[5] P. Bak, K. Chen, J. Scheinkman, and M. Woodford. Aggregate fluctuations from independent sectoral shocks: Self organized criticality in a model of production and inventory dynamics. *Ricerche Economiche*, 41:3–30, 1993. → pages 86

[6] L. Ball, N. G. Mankiw, and R. Reis. Monetary policy for inattentive economies. *Journal of Monetary Economics*, 52:703–725, 2005. → pages 86

[7] A. B. Bernard, S. Redding, and P. K. Schott. Multi-product firms and product switching. NBER Working Paper 12293, National Bureau of Economic Research, Inc, 2006. → pages 29

[8] C. Broda and D. E. Weinstein. Product creation and destruction: Evidence and price implications. NBER Working Papers 13041, National Bureau of Economic Research, Inc, 2007. → pages vii, 29, 37

[9] R. J. Caballero and A. B. Jaffe. How high are the giants' shoulders: An empirical assessment of knowledge spillovers and creative destruction in a

model of economic growth. NBER Working Papers 4370, National Bureau of Economic Research, Inc, 1993. → pages 20

[10] L. M. B. Cabral and J. Mata. On the evolution of the firm size distribution: Facts and theory. *American Economic Review*, 93:1075–1090, 2003. → pages 7

[11] J. Cai and N. Li. Knowledge linkages and multi-product firm innovations. Working papers, University of New South Wales, 2010. → pages 6, 34

[12] C. D. Carroll. The epidemiology of macroecnomic expectations. NBER Working Paper 8695, National Bureau of Economic Research, Inc, 2001. → pages 86, 88

[13] A. Clauset, C. R. Shalizi, and M. E. J. Newman. Power-law distributions in empirical data. *E-print*, page arXiv:0706.1062v2, 2007. → pages 132

[14] G. L. Clementi and H. A. Hopenhayn. A theory of financing constraints and firm dynamics. *Quaterly Journal of Economics*, 121:229–265, 2006. → pages 7

[15] D. Comin and M. Gertler. Medium term business cycles. *American Economic Review*, 96:523–551, 2006. → pages 85, 97

[16] D. Comin and S. Mulani. Diverging trends in aggregate and firm volatility. *Review of Economics and Statistics*, 88:374–383, 2006. → pages 19, 20, 85, 86, 97

[17] T. Conley and B. Dupor. A spatial analysis of sectoral complementarity. *Journal of Political Economy*, 111:311–352, 2003. → pages 86

[18] T. F. Cooley and V. Quadini. Financial markets and firm dynamics. *American Economic Review*, 91:1286–1310, 2001. → pages 7

[19] S. J. Davis, J. Haltiwanger, R. Jarmin, and J. Miranda. Volatility and dispersion in business growth rates: Publicly traded versus privately held firms. Technical report, National Bureau of Economic Research, Inc, 2006. → pages ix, 19, 85, 97, 103, 104

[20] A. K. Dixit and J. E. Stiglitz. Monopolistic competition and optimum product diversity. *American Economic Review*, 67:297–308, 1977. → pages 10

[21] E. Duguet and M. MacGarvie. How well do patent citations measure flows of technology? evidence from french innovation surveys. *Economics of Innovation and New Technology*, 14:375–393, 2005. → pages 11, 23

[22] T. Dunne, M. J. Roberts, and L. Samuelson. The growth and failure of u.s. manufacturing plants. *Quarterly Journal of Economics*, 104:671–698, 1989. → pages 6

[23] W. Dupor. Aggregation and irrelevance in multi-sector models. *Journal of Monetary Economics*, 43:391–409, 1999. → pages 86

[24] S. Durlauf. Non-ergodic economic growth. *Review of Economic Studies*, 60: 349–366, 1993. → pages 86

[25] R. Ericson and A. Pakes. Marcov perfect industry dynamics: A framework for empirical analysis. *Review of Economic Studies*, 62:53–82, 1995. → pages 6

[26] D. S. Evans. The relationship between firm growth, size and age: Estimates for 100 manufacturing industries. *Journal of Industrial Economics*, 35:567–581, 1987. → pages 6

[27] X. Gabaix. Power laws in economics and finance. NBER Working Paper 14299, National Bureau of Economic Research, Inc, 2008. → pages 30, 86, 96

[28] X. Gabaix. The granular origins of aggregate fluctuations. Nber working paper, National Bureau of Economic Research, Inc, 2009. → pages 85, 86, 93

[29] C. M. Goldie. Implicit renewal theory and tails of solutions of random equations. *The Annals of Applied Probability*, 1(1):126–166, 1991. → pages 18

[30] B. H. Hall. The relationship between firm size and firm growth in the us manufacturing sector. *Journal of Industrial Economics*, 35:583–605, 1987. → pages 6

[31] B. H. Hall, A. B. Jaffe, and M. Trajtenberg. The nber patent citation data file: Lessons, insights and methodological tools. NBER Working Paper 8498, National Bureau of Economic Research, Inc, 2001. → pages 22, 93

[32] G. D. Hansen and E. C. Prescott. Capacity constraints, asymmetries, and the business cycle. *Review of Economic Dynamics*, 8(4):850–865, October 2005. → pages 99

[33] E. Helpman, M. J. Melitz, and S. R. Yeaple. Export versus *FDI* with heterogenous firms. *American Economic Review*, 94:300–316, 2004. → pages 6, 29, 111, 113

[34] P. Holme and B. J. Kim. Growing scale-free networks with tunable clustering. *Physical Review E*, 65:026107, 2002. → pages 57

[35] H. A. Hopenhayn. Entry, exit, and firm dynamics in long run equilibrium. *Econometrica*, 60:1127–1150, 1992. → pages 6

[36] M. Horvath. Cyclicality and sectoral linkages: Aggregate fluctuations from sectoral shocks. *Review of Economic Dynamics*, 1:781–808, 1998. → pages 86

[37] M. Horvath. Sectoral shocks and aggregate fluctuations. *Journal of Monetary Economics*, 45:69–106, 2000. → pages 86

[38] M. Jackson and B. W. Rogers. Meeting strangers and friends of friends: How random are socially generated networks? *American Economic Review*, 97: 890–915, 2007. → pages 55, 57, 72, 90, 126, 127

[39] A. B. Jaffe, M. Trajtenberg, and M. S. Fogarty. Knowledge spillovers and patent citations: Evidence from a survey of inventors. *American Economic Review*, 90:215–218, 2000. → pages 23

[40] A. H. Jessen and T. Mikosch. Regularly varying functions. *Publications de L'institut Mathematique*, 94:171–192, 2006. → pages 5, 30

[41] R. Joharia, S. Mannorb, and J. N. Tsitsiklis. A contract-based model for directed network formation. *Games and Economic Behavior*, 56:201–224, 2006. → pages 56

[42] B. Jovanovic. Micro shocks and aggregate risk. *Quarterly Journal of Economics*, 102:395–409, 1987. → pages 86

[43] B. Jovanovic. The diversification of production. *Brookings Papers on Economic Activity, Microeconomics*, 1993(1):197–247, 1993. → pages 6, 29

[44] H. Kesten. Random difference equations and renewal theory for products of random matrices. *Acta Mathematica*, 131:207–248, 1973. → pages 3, 18, 32, 57, 66

[45] J. M. Kleinberg. Authoritative sources in a hyperlinked environment. *Journal of Association for Computing Machinery*, 46(5):604–632, 1999. ISSN 0004-5411. → pages 4

[46] S. Klepper. Entry, exit, growth, and innovation over the product life cycle. *American Economic Review*, 86(3):562–83, 1996. → pages 6

[47] S. Klepper and P. Thompson. Submarkets and the evolution of market structure. *Rand Journal of Economics*, 34:862–888, 2007. → pages 6, 7

[48] T. J. Klette and S. Kortum. Innovating firms and aggregate innovation. *Journal of Political Economy*, 112:986–1018, 2004. → pages 1, 2, 3, 6, 7, 11

[49] T. J. Klette and A. Raknerud. How and why do firms differ? Discussion Papers 320, Research Department of Statistics Norway, 2002. → pages 6

[50] J. B. Long and C. I. Plosser. Real business cycles. *Journal of Political Economy*, 91:39–69, 1983. → pages 86, 96

[51] R. E. Lucas. On the size-distribution of business firms. *Bell Journal of Economics*, 9:508–523, 1978. → pages 6

[52] Y. Luo. Consumption dynamics under information processing constraints. *Review of Economic Dynamics*, 11:366–385, 2008. → pages 86

[53] E. G. J. Luttmer. Selection, growth, and the size distribution of firms. *Quarterly Journal of Economics*, 122:1103–1144, 2007. → pages 7, 8

[54] E. G. J. Luttmer. On the mechanics of firm growth. Working Paper 657, Federal Reserve Bank Minneapolis, 2008. → pages 6, 7, 8

[55] S. A. M. M. L. Goldstein and G. G. Yen. Problems with fitting to the power-law distribution. *THE EUROPEAN PHYSICAL JOURNAL B - CONDENSED MATTER AND COMPLEX SYSTEMS*, 41(2):255–258, 2004. → pages 132

[56] N. G. Mankiw and R. Reis. Sticky information in general equilibrium. *Journal of the European Economic Association*, 5(2-3):603–613, 04-05 2007. → pages 86

[57] M. J. Melitz. The impact of trade on aggregate industry productivity and intra-industry reallocations. *Econometrica*, 71:1695–1725, 2003. → pages 84, 90, 99

[58] G. Moscarini. Limited information capacity as a source of inertia. *Journal of Economic Dynamics and Control*, 28:2003–2035, 2004. → pages 86

[59] M. E. J. Newman. The structure and function of complex networks. *SIAM Review*, 45:167–256, 2003. → pages 69, 88, 92, 127

[60] M. Nirei. Threshold behavior and aggregate fluctuation. *Journal of Economic Theory*, 127:309–322, 2006. → pages 86

[61] R. Reis. Inattentive producers. *Review of Economic Studies*, 73:793–821, 2006. → pages 86, 93

[62] C. Rosell and A. Agrawal. University patenting: Estimating the diminishing breadth of knowledge diffusion and consumption. NBER Working Papers 12640, National Bureau of Economic Research, Inc, 2006. → pages 20

[63] E. Rossi-Hansberg and M. L. J. Wright. Establishment size dynamics in the aggregate economy. *American Economic Review*, 97:1639–1666, 2007. → pages 6

[64] M. Seker. A structural model of establishment and industry evolution : evidence from chile. Policy Research Working Paper Series 4947, The World Bank, June 2009. → pages 6

[65] C. A. Sims. Implications of rational inattention. *Journal of Monetary Economics*, 50:665–690, 2003. → pages 86

[66] R. W. Sinnott. Virtues of the haversine. *Sky and Telescope*, 68:159, 1984. → pages 27

[67] M. L. Streitweiser. The extent and nature of establishment level diversification in sixteen us industries. *Journal of Law and Economics*, 34:503–534, 1991. → pages 29

[68] J. Sutton. Gibrat's legacy. *Journal of Economics Literature*, 35:40–59, 1997. → pages 6

[69] A. Vazquez. Growing network with local rules: Preferential attachment, clustering hierarchy, and degree correlations. *Physical Review E*, 67(056104), 2003. → pages 57

[70] M. Woodford. Imperfect common knowledge and the effects of monetary policy. NBER Working Paper 8673, National Bureau of Economic Research, Inc, 2003. → pages 86

[71] A. Young. Labor's share fluctuations, biased technical change, and the business cycle. *Review of Economic Dynamics*, 7(4):916–931, October 2004. → pages 99

Appendix A

Robustness of Sectoral Firm Size Heterogeneity

This appendix introduces a method for estimating sectoral firm size heterogeneity and the data sets used for this purpose. It then shows that sectoral firm size heterogeneity varies little when firm size is measured with different proxies. As such, sectoral firm size heterogeneity is stable over time in a specific country and also highly correlated across industrialized countries.

Helpman, Melitz, and Yeaple (33) show that there are two ways to estimate the commonly used measure of firm size heterogeneity defined by the variance of log scale firm size. For a *Pareto* distributed variable X, one method is to calculate the standard deviation of log(X) which is the reciprocal of . The other method is to estimate by OLS:

$$log(Pr(X > x)) = -\mu log(x) + c$$

and use $\frac{1}{\mu}$. Theoretically, these two methods give the same estimation. In this paper, I estimate the firm size heterogeneity measure with the second method.

The data sets used here include French manufacturing firm-level data for 1997-2005 provided by Amadeus, BUREAU van DIJK, Chilean manufacturing firm-level data for 1979-1996 provided by Chile Instituto Nacional de Estadistica, and 'Industry Statistics by Employment Size' provided by the U.S. Economic Census 1997 and 2002. The French data set has information for every firm; the Chilean data set includes only firms with more than ten employees[1]. Only the U.S. data

[1]If a firm size distribution measured by the number of employees follows a *Pareto* distribution with scale parameter μ, this truncation does not affect the estimation of μ, because a *Pareto* distribution has the special feature that when the distribution is truncated from the left, the rest of the distribution on the right tail is still a *Pareto* distribution with the same scale parameter, except that the new distribution starts with a higher minimum level x_m.

set gives the number of firms for ten employment size categories: 1 to 4 workers, 5 to 9 workers, 10 to 19 workers, 20 to 49 workers, 50 to 99 workers, 100 to 249 workers, 250 to 499 workers, 500 to 999 workers, 1000 to 2499 workers, and 2500 or more workers. From these numbers, I can determine the rank of the firms with 1, 5, 10, 20, 50, 100, 250, 500, 1000, 2500 employees in their six-digit NAICS industry.

To show the proxy robustness of firm size heterogeneity with different firm size proxies, it is appropriate to use the data from France and Chile because they both have more than one proxy for firm size. Number of employees, operational turnover, sales, and value added are the alternative proxies for firm size. The U.S. data set only has the number of employees as a proxy for firm size. Sdlnl, sdlny, sdlns and sdlnva are abbreviations for standard deviation of the log(number of employees), standard deviation of log(operational turnover), standard deviation of log(sales) and standard deviation of log(value added), respectively. In the French data set (Table 1.9), these four measures for 81 four-digit NAICS sectors are highly correlated, with a correlation coefficient greater than 0.9 in all years and for all combinations of variables. In the Chilean data set (Table 1.10), the correlation coefficients between sdlns, sdlny and sdlnva are as high as those observed in the French data set, but those between sdlnl and the other three variables are lower and range between 0.6 and 0.8. A possible reason for this discrepancy is that Chilean firm data are truncated, since only firms with more than ten employees are part of this data set.

The time persistence of firm size heterogeneity appears in the U.S., French and Chilean manufacturing sectors, though these different data sets cover different time intervals. The proxy for firm size is the number of employees in all data sets. In the U.S. data (Figure 1.12), the estimations for four-digit NAICS manufacturing industries in 1997 and 2002 exhibit a tight one-to-one relationship. In France (Figure 1.13), the estimations for four-digit NAICS manufacturing industries in 1997 and 2005 also exhibit an almost perfect one-to-one pattern[2]. The Chilean data set

[2]The outlier 3122 represents the tobacco manufacturing sector. Its heterogeneity measure drops from 3.1 in 1997 to 0.9 in 2005. There were important policy changes in this sector during this eight-year period, which might cause the significant change in firm size heterogeneity. In 2001, Brussels passed a law, banning mass-media advertising of tobacco and requiring large warning labels on cigarette packages. To discourage potential new smokers, governments throughout Europe increased

(Figure 1.14) has the longest time range: 20 years. There also, the estimations for four-digit ISIC manufacturing industries in 1979 and 1996 roughly follow a one-to-one relationship. Note that Chile experienced some economic reforms during this period. The outliers typically have less than 100 establishments.

The cross-country robustness test of firm size heterogeneity is based on a comparison between French and U.S. manufacturing sectors, because they both use NAICS industry classification. There are 81 NAICS four-digit manufacturing sectors in total. The number of employees is the firm size proxy in both data sets. Figure 1.15 and Figure 1.16 show that the correlation coefficient between sdlnl, standard deviation of log(number of employees), in the two countries is 0.74 in 1997 and 0.72 in 2002. This result corroborates a similar result in Helpman, Melitz, and Yeaple (33). They find that, although the U.S. and France have different economic policies and institutions, firm size distributions for the same industries are highly correlated across countries, with a correlation coefficient of more than 0.5.

---

their cigarette taxes in 2003.

Appendix B

Detailed One-Sector Model

The consumer's utility function becomes:

$$U = \max_{\left\{x_{ik,t}\right\}} \int_0^\infty \rho^t \left[\sum_{k=1}^{K} s^k \log\left(C_t^k\right) \right] dt,$$

$$C_{k,t} = \left(\int_0^{I_t^k} \left(x_{i,t}^k\right)^{\frac{\sigma^k - 1}{\sigma^k}} di \right)^{\frac{\sigma^k}{\sigma^k - 1}}, \ k = 1, 2, ..., K.$$

s is consumer preference for goods in sector k or the share of income spent in sector k. I_t^k is the total number of varieties in sector k. σ^k is the elasticity of substitution between differentiated goods in sector k.

The firm's problem is:

$$\max_{\left\{N_{f,t}^{ij},\, M_{f,t}^{i}\right\},\ i,\ j \in \{1,2,...,K\}} V(z_{f,t}) = \sum_{i=1}^{K} \frac{s^i P_t C_t}{\sigma_i} \frac{z_{f,t}^i}{I_t^i} + \frac{\rho C_t P_t}{C_{t+1} P_{t+1}} E[V(z_{f,t+1})] - \sum_{i=1}^{K} \sum_{j=1}^{K} \frac{\left(N_{f,t}^{ij} + M_{f,t}^{ij}\right)}{\bar{Z}_t^i}$$

subject to

$$z_{f,t+1} = z_{f,t} + \Delta z_{f,t}^N + \Delta z_{f,t}^M \qquad \text{(B..1)}$$

$$\frac{\Delta z_{f,t}^{N,i}}{z_{f,t}^i} = \sum_j \left[\frac{A_N^{ij} \left(N_{f,t}^{ij}\right)^{\alpha} \left(z_{f,t}^j\right)^{1-\alpha}}{z_{f,t}^i} + \varepsilon_{f,t}^{N,ij} \right], \ i,\ j \in \{1, 2, ..., K\} \qquad \text{(B..2)}$$

$$\frac{\Delta Z_{f,t}^{M,i}}{\bar{Z}_t^i} = \frac{A_M^{ij}\left(M_{f,t}^{ij}\right)^{\alpha}\left(\bar{Z}_t^j\right)^{1-\alpha}}{\bar{Z}_t^i} + \varepsilon_{f,t}^{M,i},\ i \in \{1,2,...,K\} \tag{B..3}$$

where $\Delta z_{f,t}^N$ and $\Delta z_{f,t}^M$ are K dimension vectors. The i^{th} element of $\Delta z_{f,t}^N$ ($\Delta z_{f,t}^M$), $\Delta z_{f,t}^{N,i}$ ($\Delta z_{f,t}^{M,i}$) is the number of innovated (imitated) new goods in sector i by firm f at time t. $\left\{\varepsilon_{f,t}^{N,ij}, \varepsilon_{f,t}^{M,i}\right\}$ are i.i.d. across firms and time.

An educated guess for the firm value function is $V(z_{f,t}^i) = \sum_{i=1}^K v_t^i \frac{z_{f,t}^i}{I_{i,t}} + u_t$. The first order conditions and Bellman equation can be written as:

$$N_{f,t}^{ij} = \left(\frac{A_N^{ij}\alpha v_t^i \rho_t^{ij}}{M_F}\right)^{\frac{1}{1-\alpha}} z_{f,t}^j,\ i,\ j \in \{1,2,...,K\} \tag{B..4}$$

$$M_{f,t}^{ij} = \left(\frac{A_M^{ij}\beta v_t^i \rho_t^{ij}}{M_F}\right)^{\frac{1}{1-\alpha}} \bar{Z}_t^j,\ i \in \{1,2,...,K\} \tag{B..5}$$

$$\begin{aligned} v_t^j &= \frac{s^j P_t C_t}{\sigma^j} - M_F \sum_{i=1}^K \left(\frac{A_N^{ij}\alpha v_{t+1}^i \rho_t^{ij}}{M_F}\right)^{\frac{1}{1-\alpha}} && \text{(B..6)} \\ &+ \left[\sum_{i=1}^K \frac{I_t^j}{I_t^i}\rho_t^{ij} v_{t+1}^i A_N^{ij}\left(\frac{A_N^{ij}\alpha v_{t+1}^i \rho_t^{ij}}{M_F}\right)^{\frac{\alpha}{1-\alpha}} + v_{t+1}^j\right], && \text{(B..7)} \\ i,\ j &\in \{1,2,...,K\}, && \text{(B..8)} \end{aligned}$$

where

$$\rho_t^{ij} = \frac{\rho I_t^j C_t P_t}{I_{t+1}^i C_{t+1} P_{t+1}}.$$

In (B..4) and (B..5), the input in each type of R&D is proportional to the knowledge capital input. In (B..6), the marginal value of one product in sector j, v_j ,depends on its current profit in sector j plus its contribution to future innovation in all K sectors.

The firm size dynamic in sector i is:

$$
\begin{aligned}
z^i_{f,t+1} &= \frac{I^i_t}{I^i_{t+1}}\left(1+A_N^{ii\frac{1}{1-\alpha}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\alpha}{1-\alpha}}+\varepsilon^{N,ii}_{f,t}\right) z^i_{f,t}+ \\
& \frac{I^j_t}{I^i_{t+1}}\sum_{j\neq i}\left(A_N^{ij\frac{1}{1-\alpha}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\alpha}{1-\alpha}}+\varepsilon^{N,ij}_{f,t}\right) z^j_{f,t}+\sum_{j=1}^{K}\frac{I^j_t}{I^i_{t+1}}\left(A_M^{ij\frac{1}{1-\beta}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\beta}{1-\beta}}+\varepsilon^{M,i}_{f,t}\right).
\end{aligned}
$$

The firm size dynamics in all K sectors are summarized by

$$
z_{f,t+1} = R_{f,t} z_{f,t} + L_{f,t}, \tag{B..9}
$$

where

$$
R^{ij}_{f,t+1} \equiv \frac{I^i_t}{I^i_{t+1}}\left(\left(A^{ij}_N\right)^{\frac{1}{1-\alpha}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\alpha}{1-\alpha}}\right)+\varepsilon^{N,ij}_{f,t}, j\neq i,\ i,\ j\in\{1,2,...,K\}
$$

$$
R^{ii}_{f,t+1} \equiv \frac{I^i_t}{I^i_{t+1}}\left(1+\left(A^{ii}_N\right)^{\frac{1}{1-\alpha}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\alpha}{1-\alpha}}\right)+\varepsilon^{N,ii}_{f,t},\ i\in\{1,2,...,K\}
$$

$$
L^i_{f,t+1} \equiv \sum_{j=1}^{K}\frac{I^j_t}{I^i_{t+1}}\left(A^{ij}_M\right)^{\frac{1}{1-\alpha}}\left[\frac{\alpha v^i_t \rho^{ij}_t}{M_F}\right]^{\frac{\alpha}{1-\alpha}}+\varepsilon^{M,i}_{f,t}.
$$

B..1 General Equilibrium

In general equilibrium, the sectoral marginal firm value, $\{v^i\}$; the number of goods growth rate g; the relative size between sectors $\left\{\frac{I^j}{I^i}\right\}$ $i,\ j=1,2,...,K$; the total consumption expenditure PC; and the number of firms M_F are solved by the following equations:

$$\left(1-\frac{\rho}{1+g}\right)\frac{v_j}{M_F}=\frac{s^jPC}{\sigma^jM_F}+\frac{1-\alpha}{\alpha}\sum_{i=1}^{K}\frac{I^j}{I^i}\left(\frac{A_N^{ij}\rho\alpha v^i}{(1+g)M_F}\right)^{\frac{1}{1-\alpha}},\ i,\ j\in\{1,2,...,K\}; \tag{B..10}$$

$$g=g^i\equiv\sum_{j=1}^{K}\left[\frac{\rho\alpha v^i}{(1+g)M_F}\right]^{\frac{\alpha}{1-\alpha}}\left[\left(A_N^{ij}\right)^{\frac{1}{1-\alpha}}+\left(A_M^{ij}\right)^{\frac{1}{1-\alpha}}\right]\frac{I^j}{I^i},\ i,\ j\in\{1,2,...,K\}; \tag{B..11}$$

$$F=u=\frac{1-\alpha}{\alpha\left(1-\frac{\rho}{1+g}\right)}\sum_{i=1}^{K}\sum_{j=1}^{K}\frac{I^j}{I^i}\left(\frac{A_M^{ij}\rho\alpha v^i}{(1+g)M_F}\right)^{\frac{1}{1-\alpha}}; \tag{B..12}$$

$$PC=L+\sum_{i=1}^{K}\frac{s^iPC}{\sigma^i}-\frac{\rho\alpha g}{1+g}\sum_{i=1}^{K}v^i. \tag{B..13}$$

In the equations above, i represents a knowledge learner sector, j represents a knowledge giver sector. Note that the number of goods in every sector is growing at the same speed because inter-sector knowledge spillovers keep all sectors on the same growing track. If one sector i had been growing more slowly than other sectors for a lengthy period, its number of goods would be very small relative to other sectors. The cross-sector knowledge spillovers would push up g_i to infinity through a huge relative sector size, $\frac{I^j}{I^i}$, in (B..11), until g^i is equal to the common growth rate.

Equation (B..10) shows that the marginal value of sector j knowledge v^j depends not only on the discounted future profit from the self-sector but also its contribution to the knowledge production in all related sectors i such that $A^{ij}>0$. If sector j gives intensive knowledge outflows to many other sectors, sector j attracts more R&D investment than others because of its higher marginal value of knowledge v_j.

Consider the special case when $A_M\equiv\gamma A_N\equiv\gamma A$. The asymmetry of knowledge diffusion matrix A indirectly shapes R&D resource allocation across sector $\left\{\frac{I^j}{I^i}\right\}$ by the relative marginal value of sectoral knowledge $\left\{\frac{I^j}{I^i}\right\}$. Again v_j is the private return of knowledge accumulation in sector j, a stronger IPRP or a lower γ increases v^j for all sector j. However, from the second part of (B..10), the increment is higher if sector j contributes more knowledge to other sectors; because the $\left\{v^i\right\}$ increments in all knowledge learner sector i add up to v^j at the rate of $\left(v^i\right)^{\frac{1}{1-\alpha}}$.

In the one-sector model, the optimal IPRP $\frac{A_M}{A_N} \equiv \gamma$ that maximizes economic growth rate depends on the trade-off between a higher imitation rate and a lower private return of knowledge v. In the multi-sector model, there is another factor to consider: the relative sector size $\left\{\frac{I^c}{I^p}\right\}$, where c represents a center sector that contributes intensive knowledge spillovers to other sectors and p represents a peripheral sector that barely generates externality. Since a higher γ hurts the private knowledge return v^c of center sectors more than that of peripheral sectors v^p, firms allocate a relatively smaller share of R&D resources to the center sector, therefore relative sector size in terms of number of goods $\left\{\frac{I^c}{I^p}\right\}$ becomes smaller. As a result, the smaller relative size of a knowledge contributing sector to a knowledge learner sector hurts the economic growth rate as illustrated in (B..11).

Appendix C

Robustness Checks

This appendix provides robustness check to the one-sector model's implications with alternative citation data sets. The first robustness check is done with random simulated citations. The second robustness check is done with all G7 country citations, which include more than 90% of all patents in NBER Patent Database, while US patents only account for about 50% of all patents.

C..1 Random Citation Data

A concern to the one-sector model's third implication is that cross-sector citations should be fewer in a sector with more heterogeneous firm size distribution, even when cross-firm knowledge diffusion is equally complete and instant in each sector. The reason is that there are more big firms in a heterogeneous sector and a big firm is more likely to cite its own patent simply because it has more patent stock available to be cited.

I simulate such random citation data sets to mimics an environment in which information is complete in every sector and compare them with the real citation data. Then I show that the real citation data set exhibits significantly larger cross-sector differences in knowledge diffusion than the random citation data sets.

In the random citation data set, the citing patent is kept the same as in the real citation data, but the cited patent is randomly assigned. Every existing patent in the same sector has an equal chance to be cited, regardless of the distance and other characteristics of the citing firm and the cited firm. I simulated 100 such random citation data sets. The values reported in Figure 1.17 are the median of these 100 data sets' results.

Figure 1.17 shows that cross-firm citations account for 95% to 99.9% of total random citations across sectors. Although it seems that a sector with heterogeneous firm size distribution has a lesser percentage of cross-firm citations than other sectors, the 5% cross-sector gap is trivial as compared with the 40% gap in the real citation data set (see Figure 1.10).

I run the same regressions as those in Table 1.1 using the 100 random citation data sets and report the results in Table 1.11. The coefficients and the robust standard errors reported are the median value of the 100 regression results. Compared with Table 1.1, the regression results using random citation data show that distance does not delay knowledge diffusion half as much as it does in real citation data; bigger firms do not cite outside patents faster than smaller firms at all; and citation time lag is slightly positively correlated with sectoral firm size heterogeneity, but the regression coefficients are much smaller than those in Table 1.1. Note that two factors still affect citation time lag in a similar magnitude to Table 1.1. The citation time lag is smaller when the cited patent is owned by a bigger firm and when the sector has a larger patent stock.

Note that these 'random citations' are not purely random, because the knowledge receiver, the citing firm, is still the same as in the real citation data, only the knowledge giver, the cited firm, is random. This is why cross-sector differences in knowledge spillovers do not disappear completely in the simulated random citation data sets.

Above all, the cross-sector differences in knowledge diffusion that exist in the real citation data set are dramatically smaller in random citation data sets, where knowledge spillovers are equally complete and instant in all sectors by construction.

C..2 G7 Country Citation Data

The G7[1] country citation data set also supports the one-sector model's implications. All estimation methods used are the same as those in section 3.

With similar results as Figure 1.7, Figure 1.18 and Figure 1.19 show that innovation rate is independent of firm size and imitation rate declines with firm size

[1]Canada, France, Germany, Italy, Japan, U.K. and U.S.

in the G7 country citation data set. Moreover, the scale-independency of imitation rate is negatively related to sectoral firm size heterogeneity.

In line with Figure 1.9, Figure 1.20 supports the finding that the sectoral firm size heterogeneity is negatively related to the ratio between the imitation's contribution to gross growth rate and the innovation risk's volatility in the larger data set with G7 country citation data.

Figure 1.21 shows the same negative relation between the cross-firm knowledge diffusion abundance and the sectoral firm size heterogeneity as Figure 1.10.

I run the same regressions as those in Table 1.1 using the G7 citation data and report the results in Table 1.14. The sectoral firm size heterogeneity is statistically significantly positively correlated with citation time lag in the regression No. 3 with country pair-industry fixed effect, but I find the same coefficient is not statistically significant in the regression No. 2 and barely significant at 10% level in the regression No. 1. My explanation is that international citations may involve more country pair-industry specific unobserved variables that are correlated with the sectoral firm size heterogeneity measure and the citation time lag at the same time, therefore firm size heterogeneity measure is not significant in the first two regressions.

Appendix D

New Imitation Production Function

In this section, I extend the basic one-sector model to allow firms to combine private and public knowledge in imitation, while using only private knowledge in innovation. Everything else remains the same, except that the imitated new goods production function becomes

$$E\left(\Delta z_{f,t}^{N}\right) = A_M M_{f,t}^{\beta} \left(\gamma z_{f,t} + (1-\gamma)\bar{Z}_t\right)^{(1-\beta)},$$

where γ is private knowledge's share in the combined knowledge pool. This imitation function implies that a firm's past R&D experience helps to absorb current public knowledge. Put another way, positive sorting in a firm's social network denotes that a larger firm can expects to learn more from the public knowledge pool, because firms of similar size are more likely to be connected. γ reflects the significance of positive sorting in the social network. These ideas are in line with the facts in patent citation data: when citing other firms' patents, firms with more patent stock tend to cite newer existing patents, cite larger firms' patents, and cite more diversified sources than smaller firms.

The firm's problem becomes

$$\max_{N_{f,t},\, M_{f,t}} V(z_{f,t}) = \frac{P_t Y_t}{\sigma} \frac{z_{f,t}}{I_t} + \rho E[V(z_{f,t+1})] - \frac{N_{f,t} + M_{f,t}}{\bar{Z}_t}$$

subject to

$$z_{f,t+1} = z_{f,t} + \Delta z_{f,t}^{N} + \Delta z_{f,t}^{M}$$

$$\frac{\Delta z^N_{f,t}}{z_{f,t}} = \frac{A_N N^{\alpha}_{f,t} z^{1-\alpha}_{f,t}}{z_{f,t}} + \varepsilon^n_{f,t}$$

$$\frac{\Delta z^M_{f,t}}{\bar{Z}_t} = \frac{A_M M^{\beta}_{f,t} \left(\gamma z_{f,t} + (1-\gamma)\bar{Z}_t\right)^{(1-\beta)}}{\bar{Z}_t} + \varepsilon^m_{f,t}$$

$$N_{f,t} = \left(\frac{A_N \alpha v \rho I_t}{I_{t+1} M_F}\right)^{\frac{1}{1-\alpha}} z_{f,t}$$

$$M_{f,t} = \left(\frac{A_M \beta (1-\gamma) v \rho I_t}{I_{t+1} M_F}\right)^{\frac{1}{1-\beta}} \left(\gamma z_{f,t} + (1-\gamma)\bar{Z}_t\right)$$

$$v = \frac{P_t Y_t}{\sigma} + \frac{\rho v I_t}{I_{t+1}} \left[1 + (1-\alpha) A_N \left(\frac{A_N \alpha v \rho I_t}{I_{t+1} M_F}\right)^{\frac{\alpha}{1-\alpha}} + \gamma (1-\beta) A_M \left(\frac{A_M \beta (1-\gamma) v \rho I_t}{I_{t+1} M_F}\right)^{\frac{\beta}{1-\beta}}\right]$$

Larger private knowledge's share in imitation γ boosts marginal firm value v, because future return on imitation also relies on current firm size. In the social network environment, a larger size today wins the firm a better peer to imitate in the future.

Higher γ also induces larger firm size heterogeneity in the sector. When private knowledge is more important in imitation or the social network is more positively assorted, sectoral firm size heterogeneity is larger for given productivity of innovation and imitation (A_N and A_M). In other words, rising γ incurs the same impact on firm size heterogeneity as rising innovation productivity A_N or decreasing imitation productivity A_M.

Appendix E

Related Questions

The following questions are closely related to the questions asked in this paper and I would like to study them further in future work. The subsequent paragraphs provide preliminary answers.

Question 1: Why do larger firms give fewer citations to peers? First, generally speaking, larger firms own better technology than average firms in the sector, therefore they are picky when using other's knowledge. The citation pattern differences across firms confirm this hypothesis. Compared with smaller firms, larger firms cite newer patents and patents owned by other larger firms. Second, larger firms tend to enter frontier subclasses in the sector, where there are only a few other competitors, which also means there are fewer targets to learn from within the subclass. Third, from a network perspective, all subclasses are connected by cross-subclass citations. These frontier subclasses locate at the periphery of the network, having fewer cross-subclass links than those subclasses at the center of the network. As a result, large firms in these frontier subclasses also make fewer cross-subclass citations. In summary, larger firms give fewer citations to others not because they do not desire to learn from others, but because not many sufficiently good outside patents are available.

Question 2: Why do larger firms grow more slowly? Does quality adjusted firm growth rate also decline with firm size? It is easier to understand firm growth rate by looking at extensive margin and intensive margin separately. g_{ex} is the growth rate attributable to patent applications in new subclasses. g_{in} is the growth rate stimulated by patent applications in existing subclasses. $Appno_t^{New}$ is the number of patent applications in new subclasses at time t. PS_{t-1} is the total number of patent stock by time t-1. $Appno_t$ is the total number of patent applications at time t. Firm growth rate drops in both margins, and it drops faster in intensive margin than extensive margin.

$$g_{ex} = ((Appno_t^{New})/(PS_{t-1}))$$

$$g_{in} = ((Appno_t - Appno_t^{New})/(PS_{t-1}))$$

There are two ways to look at the decreasing intensive growth margin. On one hand, products of the same firm may be closer substitutes than products of different firms. Since a firm's new product becomes a close competitor of its previous products in the same subclass, the return from one additional product in an existing subclass decreases as a firm accumulates more products in the same subclass.

On the other hand, the gain from learning is smaller when a firm owns more knowledge in a small subclass. The network structure among subclasses helps to analyze the declining intensive margin. When a firm dynamically expands across subclasses, it normally starts from the center of the network and gradually enters other linked subclasses farther away from the center. Only top firms enter frontier subclasses at the periphery of the network, looking for higher profit margin and weaker competition. The isolated network location of the periphery subclasses is a natural obstacle blocking small firms without related knowledge from entering. Since the center subclasses are well connected with other subclasses, smaller firms, which usually start from the center, have access to many potential new subclasses. In contrast, when larger firms approach the edge of the network, the network becomes sparser. There are fewer linked new subclasses that they can potentially enter. Quality adjusted growth rate is measured by the citation-weighted number of patents, instead of a simple patent count. When adjusted by the number of inward citations, a larger firm's growth rate drops even faster. The reason is that the number of inward citations per patent decreases with firm size in both extensive margin (number of subclasses) and intensive margin (number of patents within the subclass).

Appendix F

Technical Details

To fit into Jackson and Rogers (38)'s non-directed network environment, I ignore the direction in the citation networks and treat them as non-directed networks. For sector s at time t, the adjacency matrix becomes $\hat{M}_{s,t}(i,j) = \max\left(M_{s,t}(i,j), M_{s,t}(j,i)\right)$. Firm i and firm j are called "old friend" at time t, if i and j are connected at both t and $t-1$ ($\hat{M}_{s,t-1}(i,j) = 1$ and $\hat{M}_{s,t}(i,j) = 1$). Firm i and j are called "new friend" if firm i does not connect with firm j at time $t-1$, but i connects with j at time t ($\hat{M}_{s,t-1}(i,j) = 0$ and $\hat{M}_{s,t}(i,j) = 1$).

Conditional on firm i and j being new friends, they are called 'network-based new friend' or 'friend's friend' to each other, if there exists at least one firm $k \neq i,j$ such that k is connected with both i and j at time $t-1$ ($\hat{M}_{s,t-1}(i,:)*\hat{M}_{s,t-1}(:,j) \geq 1$, where $M_{s,t-1}(i,:)$ is the i^{th} row of matrix $\hat{M}_{s,t-1}$ and $\hat{M}_{s,t-1}(:,j)$ is the j^{th} column of matrix $\hat{M}_{s,t-1}$). Conditional on firm i and j being new friends, they are called 'random new friend' to each other, if there is no such firm k that is connected with both i and j at time $t-1$ ($\hat{M}_{s,t-1}(i,:)*\hat{M}_{s,t-1}(:,j) = 0$).

Mathematically $r_{s,t}$ is calculated as:

$$r_{s,t} = \frac{nnz\left[not\left(\hat{M}_{s,t-1}*\hat{M}_{s,t-1}\right).*not\left(\hat{M}_{s,t-1}\right).*\hat{M}_{s,t}\right]}{nnz\left[\hat{M}_{s,t-1}*\hat{M}_{s,t-1}.*not\left(\hat{M}_{s,t-1}\right).*\hat{M}_{s,t}\right]}. \tag{F..1}$$

The numerator (denominator) in (F..1) is the number of new random (friend's) friends made in sector s at time t. nnz counts the number of none zero elements. The (i,j) element in $\hat{M}_{s,t-1}*\hat{M}_{s,t-1}$ is the number of common friends that firm i and j share at time $t-1$. $not(M)$ replaces positive elements in M with zeros, and replaces zeros with ones. Therefore the (i,j) element in $not\left(\hat{M}_{s,t-1}*\hat{M}_{s,t-1}\right)$ is one, if firm i and firm j has no common friend at time $t-1$, it is zero otherwise. The (i,j) element in $not\left(\hat{M}_{s,t-1}\right).*\hat{M}_{s,t}$ is one, if firm i and j are new friends to each other, it is zero otherwise. All together, the (i,j) element in $not\left(\hat{M}_{s,t-1}*\hat{M}_{s,t-1}\right).*not\left(\hat{M}_{s,t-1}\right).*\hat{M}_{s,t}$ is positive, if firm i and j are 'random new friends', it is zero otherwise. Similarly, the (i,j) element in $\hat{M}_{s,t-1}*\hat{M}_{s,t-1}.*not\left(\hat{M}_{s,t-1}\right).*\hat{M}_{s,t}$ is positive, if firm i and j are 'network-based new friends', it is zero otherwise.

Newman (59) gives several ways to measure network clustering. Jackson and Rogers (38) examines three commonly used clustering coefficients in the literature. They are:

$$C^{TT}\left(M_{s,t}\right)=\frac{\sum_{i;j\neq i;k\neq i,j}M_{s,t}\left(i,j\right)M_{s,t}\left(j,k\right)M_{s,t}\left(k,i\right)}{\sum_{i;j\neq i;k\neq i,j}M_{s,t}\left(i,j\right)M_{s,t}\left(j,k\right)},$$

$$C\left(M_{s,t}\right)=\frac{\sum_{i;j\neq i;k\neq i,j}\hat{M}_{s,t}\left(i,j\right)\hat{M}_{s,t}\left(j,k\right)\hat{M}_{s,t}\left(k,i\right)}{\sum_{i;j\neq i;k\neq i,j}\hat{M}_{s,t}\left(i,j\right)\hat{M}_{s,t}\left(j,k\right)},$$

and

$$C^{Avg}\left(M_{s,t}\right)=\frac{1}{n}\sum_{i}\frac{\sum_{j\neq i;k\neq i,j}\hat{M}_{s,t}\left(i,j\right)\hat{M}_{s,t}\left(j,k\right)\hat{M}_{s,t}\left(k,i\right)}{\sum_{j\neq i;k\neq i,j}\hat{M}_{s,t}\left(i,j\right)\hat{M}_{s,t}\left(j,k\right)}.$$

They all measure the likelihood that two nodes are connected, conditional on these two nodes being connected with a common node. The first two definitions are the same when the network is non-directed. The third definition gives an equal weight to every node, while the first two definitions give a bigger weight to the node with more links.

To estimate $\mu^x_{s,t}$ in sector s at time t, I run OLS regress of $ln(1-F_{s,t}\left(d^x_{f,t}\right))$ on $ln(d^x_{f,t})$, where x=$\{$in,out and total$\}$, $F_{s,t}\left(d^x\right)$ is the c.d.f. of x-degree distribution in sector s at time t. $\hat{\mu}^x_{s,t}$ is equal to the absolute value of the OLS coefficient before $ln(d^x_{f,t})$. Note that standard deviation of $\left\{ln(d^x_{f,t})\right\}$ is equal to $\frac{1}{\hat{\mu}^x_{s,t}}$, which measures the heterogeneity of x-degree distribution.

Appendix G

Simulating Networks for Every Sector

In sector s' firm citation network, I identify the realization of $F_{f,t}$ and $R_{f,t}$ for node f at time t with the following steps.

(1) Identify new friend.

If $M_{s,t}(j,i) = 1$ and $M_{s,t-1}(j,i) = 0$, node j is node i's new inward friend at time t. If $M_{s,t}(i,j) = 1$ and $M_{s,t-1}(i,j) = 0$, node j is node i's new outward friend at time t. If $M_{s,t}(j,i) = 1$ and $M_{s,t-1}(j,i) = 1$, node j is node i's old inward friend at time t.

(2) Identify the source of new friend.

Suppose node j is node i's new inward friend at time t.

If there exists a node k, such that node k is an inward friend of node i, and inward or outward friend of node j at time $t-1$; then node j is node i's new inward friend introduced by inward friend k. The total number of such node j introduced by inward friend are denoted as $\Delta d^d_{f,t}$.

If there exists a node k, such that node k is an outward friend of node i, and k is inward or outward friend of node j at time $t-1$ or $j = k$; then node j is node i's new inward friend introduced by outward friend k. The total number of such node j is denoted as $\Delta d^p_{f,t}$.

If node i has both inward and outward common friends who are friend of j, then I attribute half to $\Delta d^p_{f,t}$ and half to $\Delta d^d_{f,t}$.

If node j and i do not have any type of common friend, then node j is node i's random new friend and belongs to $\Delta d^r_{f,t}$.

Similarly, new outward friend introduced by inward friend, outward friend, and random new friend $\Delta p^d_{f,t}$, $\Delta p^p_{f,t}$, and $\Delta p^r_{f,t}$ are identified.

(3) Estimate δ, the possibility to drop an old link. If $M_{s,t}(j,i) = 0$ and $M_{s,t-1}(j,i) = 1$, then the old link between i and j is dropped at time t. Denote $drop_{f,t}$ as the number of inward link dropped by firm f at time t. In the entire network, the possibility to drop an old link is $\delta = \frac{\sum_f drop_{f,t}}{\sum_f d_{f,t-1}}$.

(4) Infer the elements in random matrix $F_{f,t}$ and $R_{f,t}$. N is the total number of nodes in the network.

$$F_{f,t}^{11} = 1 - \delta + \frac{\Delta p_{f,t}^{p}}{p_{f,t}},$$

$$F_{f,t}^{12} = \frac{\Delta p_{f,t}^{d}}{d_{f,t}},$$

$$F_{f,t}^{21} = \frac{\Delta d_{f,t}^{p}}{p_{f,t}},$$

$$F_{f,t}^{22} = 1 - \delta + \frac{\Delta d_{f,t}^{d}}{d_{f,t}},$$

$$R_{f,t}^{1} = \frac{\Delta p_{f,t}^{r}}{N},$$

$$R_{ft}^{2} = \frac{\Delta d_{ft}^{r}}{N}.$$

Appendix H

Calculation Detail

The mass of just-informed firms is:

$$
\begin{aligned}
JI(k) &= \int_{A_{\min}}^{\infty} JI(k,A)\, dF(A) = \int_{A_{\min}}^{\infty} \frac{(c_t - d_t \ln(A))^k A_m^{\mu} \mu A^{d_t-\mu-1}}{k! e^{c_t}} dA \\
&= \left[- \frac{A_{\min}^{\mu} \mu (c_t - d_t \ln(A))^k A^{d_t-\mu}}{k! e^{c_t} (\mu - d_t)} \right]_{A_{\min}}^{\infty} - \int_{A_{\min}}^{\infty} \frac{d_t A_m^{\mu} \mu (c_t - d_t \ln(A))^{k-1} A^{d_t-\mu-1}}{(\mu - d_t)(k-1)! e^{c_t}} dA \\
&= \frac{\mu (c_t - d_t \ln(A_{\min}))^k A_{\min}^{d}}{(\mu - d_t) k! e^{c_t}} - \frac{d_t}{\mu - d_t} JI(k-1) \\
&= \frac{\mu}{\mu - d_t} JI(k, A_{\min}) - \frac{d_t}{\mu - d_t} JI(k-1) \\
&= \frac{\mu}{\mu - d_t} \sum_{g=0}^{k} \left(- \frac{d_t}{\mu - d_t} \right)^{k-g} JI(g, A_{\min}).
\end{aligned}
$$

The total output is:

$$
\begin{aligned}
Y_{t+k} &= \frac{1}{P_{t+k}} = \frac{(\sigma-1)\tilde{A}_{t+k}}{\sigma} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\int_{A_m}^{\infty} I(k,A)\left[A(1+\triangle)\right]^{\sigma-1} + (1-I(k,A))A^{\sigma-1}dF(A)\right\}^{\frac{1}{\sigma-1}} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\int_{A_m}^{\infty} I(k,A)A^{\sigma-1}\left[(1+\triangle)^{\sigma-1}-1\right] + A^{\sigma-1}dF(A)\right\}^{\frac{1}{\sigma-1}} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\left[(1+\triangle)^{\sigma-1}-1\right]\int_{A_m}^{\infty} I(k,A)A^{\sigma-1}dF(A) + \frac{\mu}{\mu-\sigma+1}A_m^{\sigma-1}\right\}^{\frac{1}{\sigma-1}} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\left[(1+\triangle)^{\sigma-1}-1\right]\sum_{g=0}^{k}\int_{A_m}^{\infty} JI(g,A)A^{\sigma-1}dF(A) + \frac{\mu}{\mu-\sigma+1}A_m^{\sigma-1}\right\}^{\frac{1}{\sigma-1}} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\left[(1+\triangle)^{\sigma-1}-1\right]\sum_{g=0}^{k}\int_{A_m}^{\infty} \frac{A_m^{\mu}\mu\left(c_t-d_t\ln(A)\right)^k A^{d+\sigma-\mu-2}}{k!e^{-c_t}}dF(A) + \frac{\mu}{\mu-\sigma+1}A_m^{\sigma-1}\right\}^{\frac{1}{\sigma-1}} \\
&= \frac{(\sigma-1)}{\sigma}\left\{\frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g} JI(g,A_{\min}) + \frac{\mu}{\mu-\sigma+1}\right\}^{\frac{1}{\sigma-1}} A_m.
\end{aligned}
$$

The total output of top firms whose sizes are larger than A_x is:

$$
Y_{t+k}^{A>A_x} = \left(\frac{\tilde{A}_{t+k}|A_{t+k}>A_x}{\tilde{A}_{t+k}}\right)^{\sigma} Y_{t+k} = \frac{(\sigma-1)}{\sigma}\frac{\left[\int_{A_x}^{\infty} A_{i,t+k}^{\sigma-1}dF(A_{i,t})\right]^{\frac{\sigma}{\sigma-1}}}{\int_{A_x}^{\infty} A_{i,t+k}^{\sigma-1}dF(A_{i,t})}
$$

$$
= \frac{(\sigma-1)}{\sigma}\frac{\left[\frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g}JI(g,A_x)+\frac{\mu}{\mu-\sigma+1}\right]^{\frac{\sigma}{\sigma-1}}}{\frac{(1+\triangle)^{\sigma-1}-1}{\mu-\sigma+1-d_t}\sum_{g=0}^{k}\sum_{j=0}^{g}\left(-\frac{d_t}{\mu-\sigma+1-d_t}\right)^{k-g}JI(g,A_{\min})+\frac{\mu}{\mu-\sigma+1}}\frac{A_x^{\sigma}}{A_{\min}^{\sigma-1}}
$$

Appendix I

Goodness-of-Fit Test for Power-Law Distribution

Following the Kolmogorov-Smirnov (KS) type goodness-of-fit test for in Clauset, Shalizi, and Newman (13) and M. L. Goldstein and Yen (55), I calculate KS test critical value table, assuming MLE estimation. Then compare the KS test statistics with the correspondent critical values and number of observations.

In the patent data, firm size is measured by the number of patents. Among the 42 3-digit SIC industries, the Power-law size distribution with cut-off can not be rejected at 10% probability for all but 5 sectors (Industrial organic chemistry, Plastics materials and synthetic resins, Agricultural chemicals, Motorcycles, bicycles, and parts and Miscellaneous transportation equipment). In the patent data, in-degree (out-degree) is the number of inward (outward) citations, total degree it the summation of in-degree and out-degree. Among the 42 3-digit SIC industries, the Power-law in-degree distribution with cut-off can not be rejected at 10% probability for all sectors but Miscellaneous transportation equipment; the Power-law out-degree distribution with cut-off can not be rejected at 10% probability for all sectors but Guided missiles and space vehicles and parts and Stone, clay, glass and concrete products; the Power-law total-degree distribution with cut-off can not be rejected at 10% probability for all sectors but Refrigeration and service industry machinery, Household appliances and Railroad equipment. See column 3-6 of Table 2.2. In the French manufacturing firm data, firm size is measured by number of employees. Among 21 3-digit NAICS 2002 industries, the p-value of KS statistics is greater than 10% for all sectors except Animal Food Manufacturing sector.