Title
Three Essays in Economics

Permalink
https://escholarship.org/uc/item/4p3763xq

Author
XIE, CONG

Publication Date
2015

Peer reviewed|Thesis/dissertation

UNIVERSITY OF CALIFORNIA

Los Angeles

Three Essays in Economics

A dissertation submitted in partial satisfaction

of the requirements for the degree

Doctor of Philosophy in Economics

by

Cong Xie

2015

ABSTRACT OF THE DISSERTATION

Three Essays in Economics

by

Cong Xie

Doctor of Philosophy in Economics

University of California, Los Angeles, 2015

Professor Hugo Andres Hopenhayn, Chair

The mechanism of innovation is an important question in economics. Recent release of the patent transaction data provides some new insights in understanding the innovation process.

Chapter 1 studies the patent reassignment, which is an important phenomenon in the United States. This chapter documents some basic facts about patent transactions on which patents are traded and how they are traded. I also provide evidence on the motive and impact behind these transactions. There is more strategic knowledge flow across firms associated with mergers and acquisitions.

Chapter 2 further investigates the information flow among agents in the economy and how this affects the economic growth. This chapter studies technology diffusion in a heterogeneous agent environment. Agents can improve through imitation or innovation. The benefits of imitation was determined by the economy's distribution which evolutes over the time. I estimated the model by linking patent citations to imitation process. I provide suggestive evidence on relative contribution of imitation vs. innovation to the economic growth. I also explained the non monotone relationship between patent citations and its value documented recently.

Chapter 2 develops a general method in studying the economic meaning in the cross-section distribution of economic variables. Chapter 3 applies this method to study the income wealth distribution in US. I present some evidence on the contribution of the income process to wealth inequality using a continuous time Aiyagari type model. I then estimate the

model semi/nonparametrically with the US data. The results suggest that higher moments of the income process is important in explaining the income wealth distribution.

The dissertation of Cong Xie is approved.

Simon Adrian Board

Zhipeng Liao

Pierre-Olivier Weill

Lynne Goodman Zucker

Hugo Andres Hopenhayn, Committee Chair

University of California, Los Angeles

2015

To my parents . . .

TABLE OF CONTENTS

List of Figures

List of Tables

Vita

2005-2009   B.Sc. in Economics and Finance, Minor in Mathematics

Hong Kong University of Science and Technology

Hong Kong

2009-2010   M.A. in Economics

UCLA

Los Angeles, California

2011   M.Phil. in Economics

UCLA

Los Angeles, California

CHAPTER 1

Market for firms or market for ideas

1.1 Introduction

Patent reallocation is an important phenomenon in US. Which type of patent is more often traded? How are they traded? Why might they be traded differently? Using the recent data on patent reassignment, I try to address these questions and show that the market for firms provides an alternative to the market for patents in inducing the information flow between firms.

There is a long research tradition in studying the exchange of ideas between the individual and corporations. Many studies in patent literature focused on the role of patent citations in knowledge flow. For instance, Hall, Jaffe, and Trajtenberg (2001)(HJT01) offers a general description and treatment. Recently, Galasso and Schankerman (2015)(GS14) analyzed patent citations to study how patents affect cumulative innovation.

A second strand of the patent literature directly studies the market for patents. Lamoreaux and Sokoloff (1999)(LS99) examined the assignment of patents in the late 19th century and early 20th century and found an active patent transfer market. This is in contrast to the classic view by Schumpeter that knowledge creation should be concentrated within large corporations. A recent paper by Serrano (2010)(Ser10) reviewed more summary evidence of patent transactions in the US using a new data set by US patent and trademark office (USPTO). Akcigit (2014)(ACG14) further modeled patent transactions in a macroeconomics framework where ideas were opportunities to upgrade the technology. Whether a particular patent is useful or not is random. This mismatch creates an incentive for an active trading market of patents. Both of them focused on the frequency of the trades or reassignments in

1

USPTO.

The objective of this paper is to propose that the market for firms is a third way to induce information flow between firms and is particularly important for strategic tacit knowledge. Combining a larger data set, I study how patents were transacted, namely through direct transfer or through transfer of the firm ownership. The second method, which was labeled as Mergers and Acquisitions (m&a), is often captured more by certain corporations in the news. Recent cases like Cisco buying a firm every half year fit in this scenario. This is related to the long tradition of the discussion of the boundary of firms by Williamson (Wil10), and other extensive discussions starts there.

If a patent is the perfect claim of property rights of a science innovation and its potential further development, then the direct transfer of the property rights through a market mechanism would be sufficient. Alternatively, if there are intangible assets or extra information that is either difficult to patent or contains strategic information and becomes too sensitive to patent, then acquiring the entire firm as well as maintaining the former employees would be important for transferring that intangible or strategic information. This short paper provides some evidence on the latter by studying the impact of those patent transactions through different transaction methods. Data suggest that there is more information flow associated with patents obtained by mergers and acquisitions than from direct transfers. The market for firms is important.

The empirical work focuses on several aspects. First, I characterize the pattern of the different transaction methods and how this relates to the characteristics of the patents. Different industries have different degrees of patent protection, and the importance of the intangible assets may vary. I summarize how patent class affects the method of the transaction.

Second, I compare the differences in the utilization rate of the patent before and after the transaction. If a firm were purchased through merger and acquisition for its embedded intangible assets, then the acquiring firm would exploit these intangible assets. This increase in utilization would be reflected by the change in the (internal) citation rate.

Third, I study the content in the information flow by studying the area of the patents

using a topic clustering method. This measures whether the information is closely related with the knowledge a firm previously possessed. Information flow from a different speciality might induce a change in the innovation content of the acquiring firm. Studying the potential content change then provides evidence for the existence of an information flow.

This paper is also related to the literature on the market for firms, such as the pure asset purchasing and expansions theory, e.g. the Q theory from Jovanovic (2002)(JR02), where the physical assets of the target firm matter more. Rhodes-Kropf and Robinson (2008)(RR08) studied a firm dynamic structure where mergers is a method to improve productivity. David (2012)(Dav12) further studied productivity-enhancing mergers within a general equilibrium firm dynamic model. My paper provides evidence on where the productivity enhancement comes from.

This paper is organized as follows: Section 1.2 describes the construction of the data set and provides an overview of the data. Section 1.3 examines the differences in the information flow associated pre- and post-patent transfers and provides some suggestive evidence on the motives for mergers and acquisitions. 3.7 concludes.

1.2   Data

The original data were taken from separate sources. USPTO patent assignment data [1] were used as the basis to construct the data set. Serrano (2010)(Ser10) and Kogan, Papanikolou and Seru (2011)(KPS12) provides a detailed treatment on how to process the data. This assignment data contains the assignments from 1976 to 2014, and information on the subsequent transactions after the initial assignment. The USPTO assignment data provide information on the transacting parties: assignors, the previous owner of the patent; assignees, the new owner of the patent; the date of the reassignment and the reason of the reassignment. Following Akcigit (2014)(ACG14) I classified the reasons into the following categories:

---

[1]Google bulk download

- Initial assignment

- Assignor's interest which is classified as direct transfer

- Mergers and Acquisitions

- Release

- Lease and lien

- Mortgage

- Liquidation

- Employment agreement

- Surrender and Abandon

- Trade secretes non-compete

- Corporate dissolution

- Court

- Others

Among those transaction reasons, direct transfer and merger and acquisitions were the top reasons. In the original USPTO data, patents reassigned through mergers and acquisitions make up about 1.5% of the number of patents reassigned through direct transfers. However, if we update mergers and acquisitions, the number of patents transferred by mergers and acquisitions significantly increased.

The next data source is the maintenance fee event of the USPTO. This documents the application date and some subsequent event of the patent. Most importantly, this contains the information on expiration date of the patent. The ownership of the patent is terminated by the expiration of the patent. This data set also provides information on whether the owner is a big entity or not.

SDC mergers and acquisitions data from Thompson Reuters SDC platinum provided the records of the transaction of the firms. It contains m&a data from 1976 to 2014. Inventor data from Zucker and Darby (2011)(ZDF11) and the Fleming (2011)(LDY11) data set were used for detecting the same inventor.

The fourth major data source is the citation data, and the classification data from the USPTO, national bureau of economic research (NBER) pdp project and Comets database. Information on the citations is used to measure the utilization of the patent and other citation based dependent variables. Combined with the ownership data, I differentiate the citations into external citations, internal citations on the externally obtained patents, and internal citations on the internally developed patents. All externally obtained patents were linked to their transaction date, ownership duration, and transaction format from the dynamic ownership data.

1.2.1 Dynamic ownership data

The original USPTO data provide the basis for the transaction record. M&as will give the acquiring firm of the property rights or at least access to the patents owned by the target firm. The USPTO did not pick up all those changes. On the other hand, if the target firm buys and sells patents actively, then the patent portfolio acquired by the acquiring firm through mergers and acquisitions is quite different from the initial assignment owned by the target firm. This is why both USPTO reassignment data and SDC m&a data are necessary to construct the patent portfolio. Combining these three data sets, I build the dynamic ownership data. On a patent level, this traces the life cycle and events of a patent. For a given time I can trace the patent portfolio of a firm.

1.2.2 Summary of the data

There are around 2.7 million patents issued. Although only less than 20% of the patents have ever been transacted, the reallocation activity is significant. Patents have been traded multiple times. The reallocation activity also highly correlates with the business cycle (sim-

ilar to studies found in Serrano (2010)(Ser10)). If we filter the reallocation activity with the Hodrick-Prescott (hp) filter at a frequency of 100 and compare this to the business cycle data from Eisfeldt and Rampini (2006)(ER06), we find a high correlation between the two. The three peaks reflected the three merger waves.
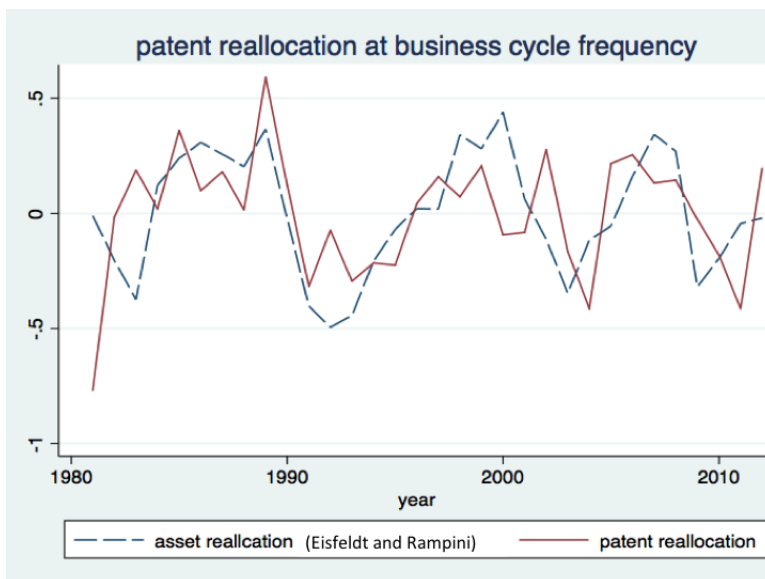
Figure 1.1: Relationship with the business cycle.[2]

Which type of patents that is transacted? Among these transactions chemicals, biology, and medicine patents account for 32% of these transactions using the Zucker and Darby scientific classification. The Hall, Jaffe and Trajchenberg measure shows a higher share of these categories, which is around 74.7%.

HJT-class	Percentage	ZD-class	
Chemical	16.1%	Biology, chemistry & medicine	16 %
Computer & communication	27.8%	Computing & information technology	16 %
Drugs & medical	6.6%		
Electrical & electric	24.2%	Semiconductor	5.3%
Mechanical	14.7%		
Others	11%		

Table 1.1: Decomposition of the transaction by class classification.

6

The patent literature often uses citations or some variants of citations as a measure of importance. More citations mean they are essential in providing guidance for future technology and patenting behavior. Non-patent referencing was used by many researchers, such as Roach and Cohen (2012)(RC13), as a measure complementing the patent citation. Traded patents are more important by these three measures.

Henderson, Jaffe and Trajtchenberg (1998)(HJT98) defined generality and originality, which measure the spread of citations across the different patent classes.[3] If citations received are equally spread out, it has a higher generality score. Originality is defined similar using citations made. It is used by Hall, Jaffe and Trajtchenberg (2001)(HJT01) and later patent studies. Traded patents show less generality and higher originality, although the difference in originality is insignificant. Regression results are listed in Table 1.2, where 70-90 cohort is an indicator variable that takes 1 if the patent was applied between 1970 and 1990. 91-00 cohort is defined similarly.

---

[3]The spread out is measured by one minus the herfindahl index of citations.

Variables	(1) Citations received	(2) citations made	(3) non-patent referencing count	(4) generality	(5) originality
Trnsacted	2.126***	3.852***	0.264***	-0.00419***	0.00157
	(0.0680)	(0.0716)	(0.0471)	(0.00115)	(0.00115)
citations received		0.0601***	0.249***	-0.000285***	0.001000***
		(0.00172)	(0.00105)	(2.77e-05)	(2.75e-05)
citations made	0.0539***		0.0130***	-2.44e-05	4.94e-05*
	(0.00155)		(0.00107)	(2.62e-05)	(2.61e-05)
nprcount	0.521***	0.0304***		0.000275***	0.000661***
	(0.00221)	(0.00250)		(4.00e-05)	(3.98e-05)
generality	-0.995***	-0.0951	0.458***		0.348***
	(0.0966)	(0.102)	(0.0668)		(0.00152)
originality	3.530***	0.194*	1.115***	0.352***	
	(0.0970)	(0.103)	(0.0672)	(0.00154)	
70-90 cohort	-13.89***	14.68***	-1.913***	0.100***	-0.0406***
	(0.138)	(0.146)	(0.0966)	(0.00236)	(0.00235)
91-00 cohort	-8.417***	8.785***	-0.445***	0.0749***	-0.0234***
	(0.129)	(0.136)	(0.0897)	(0.00219)	(0.00218)
chem	-3.454***	-0.794***	2.295***	0.0247***	0.00223
	(0.119)	(0.125)	(0.0820)	(0.00201)	(0.00200)
Computer & communication	-1.896***	5.667***	1.106***	-0.0577***	-0.0759***
	(0.112)	(0.118)	(0.0773)	(0.00189)	(0.00188)
Drugs & medical	-3.851***	4.152***	8.837***	-0.0730***	-0.0371***
	(0.129)	(0.136)	(0.0880)	(0.00218)	(0.00217)
Electrical & electric	-3.415***	2.279***	0.654***	-0.0446***	-0.0709***
	(0.114)	(0.121)	(0.0792)	(0.00194)	(0.00192)
Mechanical	-1.809***	-0.288**	-0.259***	0.00528**	-0.0140***
	(0.132)	(0.139)	(0.0912)	(0.00223)	(0.00222)
Constant	18.54***	-0.782***	0.681***	0.316***	0.402***
	(0.166)	(0.178)	(0.117)	(0.00281)	(0.00276)
Observations	373,920	373,920	373,920	373,920	373,920
R-squared	0.183	0.055	0.186	0.148	0.148

Standard errors in parentheses

*** p<0.01, ** p<0.05, * p<0.1

Table 1.2: Transacted patents.

These results suggest that trade patents are generally more important. Since a patent is an important certificate of technology, firms want to trade patents to utilize specific production technology. Those patents thus might have a narrower scope and a lower generality.

How are they traded? Second, I address the questions on whether patents are traded differently and are there any differences in the patents using different transaction methods.

Among all the transactions, patents reallocated by mergers and acquisitions accounts for 80-90% of the total reallocation, followed by direct transfers, which accounts around 10% of the total transactions. This is quite different from the view that patents are liquid and transacted on an individual level for technology-upgrade opportunities as in Akcigit (2014)(ACG14). Among all the categories, computer & communication, electrical electric patents have a higher tendency to be traded through mergers and acquisitions. Direct transfer is more appealing for chemical and drugs & medical patents. Within each class, the subclasses also show variation in these shares. Figure 1.3 depicts this within-category difference.
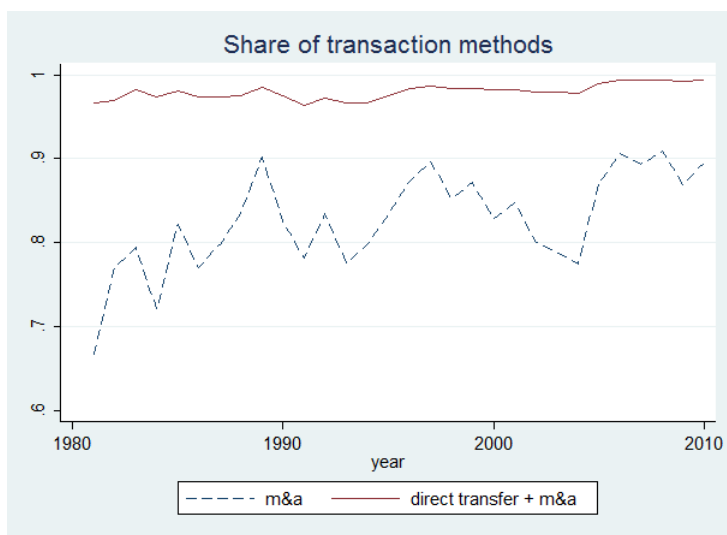
Figure 1.2: Relative shares of different transaction methods.

HJT-class	M&a percentage	Direct transfer percentage
Chemical	87.1%	11.7%
Computer & communication	89.8%	9.1%
Drugs & medical	75.2%	22.4%
Electrical & electric	90.8%	8.37%
Mechanical	85%	13.2%
Others	76.6%	20.3%

Table 1.3: Decomposition of the transaction by class classification.

9
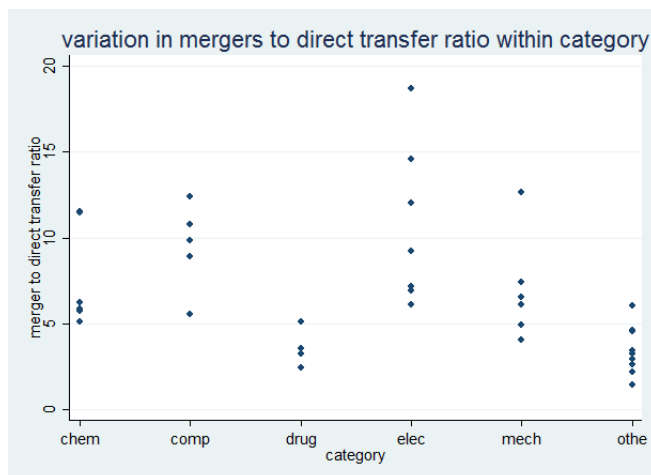
Figure 1.3: Variation of mergers to direct trade ratio within category.

Similarly if we examine citation-based measures, patents traded through mergers and acquisitions tend to have more citations received (by 83%), citations made (by 17%) and significantly lower generality and lower originality than patents trade through direct transfer. This is supported with a level of significance at 1%. The non-patent referencing did not show significant differences. Results are presented in Table 1.4.

	(1)	(2)	(3)	(4)	(5)
Variables	citations received	citations made	non patent referencing count	generality	originality
merger	2.193***	3.776***	0.256***	-0.0105***	-0.00321**
	(0.0742)	(0.0782)	(0.0513)	(0.00126)	(0.00125)
direct trade	1.208***	3.225***	0.194**	0.0152***	0.0151***
	(0.120)	(0.127)	(0.0829)	(0.00203)	(0.00202)
citations received		0.0601***	0.249***	-0.000283***	0.00100***
		(0.00172)	(0.00105)	(2.77e-05)	(2.75e-05)
citations made	0.0539***		0.0130***	-2.34e-05	4.96e-05*
	(0.00155)		(0.00107)	(2.62e-05)	(2.61e-05)
non patent referencing count	0.521***	0.0304***		0.000275***	0.000661***
	(0.00221)	(0.00250)		(4.00e-05)	(3.98e-05)
generality	-0.986***	-0.0912	0.459***		0.347***
	(0.0966)	(0.102)	(0.0668)		(0.00152)
originality	3.538***	0.195*	1.115***	0.352***	
	(0.0970)	(0.103)	(0.0672)	(0.00154)	
70-90 cohort	-13.88***	14.67***	-1.914***	0.0998***	-0.0407***
	(0.138)	(0.146)	(0.0966)	(0.00236)	(0.00235)
91-00 cohort	-8.406***	8.775***	-0.446***	0.0745***	-0.0236***
	(0.129)	(0.136)	(0.0897)	(0.00219)	(0.00218)
Chemical	-3.478***	-0.815***	2.293***	0.0252***	0.00258
	(0.119)	(0.125)	(0.0821)	(0.00201)	(0.00200)
Computer & communication	-1.925***	5.640***	1.104***	-0.0569***	-0.0754***
	(0.112)	(0.118)	(0.0773)	(0.00189)	(0.00188)
Drugs & medical	-3.857***	4.136***	8.836***	-0.0731***	-0.0372***
	(0.129)	(0.136)	(0.0880)	(0.00218)	(0.00217)
Electrical & electric	-3.444***	2.263***	0.652***	-0.0438***	-0.0702***
	(0.115)	(0.121)	(0.0793)	(0.00194)	(0.00193)
Mechanic	-1.826***	-0.305**	-0.260***	0.00558**	-0.0138***
	(0.132)	(0.139)	(0.0912)	(0.00223)	(0.00222)
Constant	18.57***	-0.733***	0.686***	0.316***	0.402***
	(0.166)	(0.178)	(0.117)	(0.00281)	(0.00276)
Observations	373,927	373,927	373,927	373,927	373,927
R-squared	0.183	0.056	0.186	0.149	0.148

Standard errors in parentheses

*** $p<0.01$, ** $p<0.05$, * $p<0.1$

Table 1.4: Difference between patents according to different transaction methods.

The average age of a patent at transaction is about 8.43 years after the application and 6.07 years after the grant. There is a significant number of the transactions took place between the application and grant period, accounting for 9.32% of the total transactions. Figure 1.4 plots the distribution of the patent age at the transaction. The age distribution

of patents transacted in the mergers and acquisitions has a fatter tail.
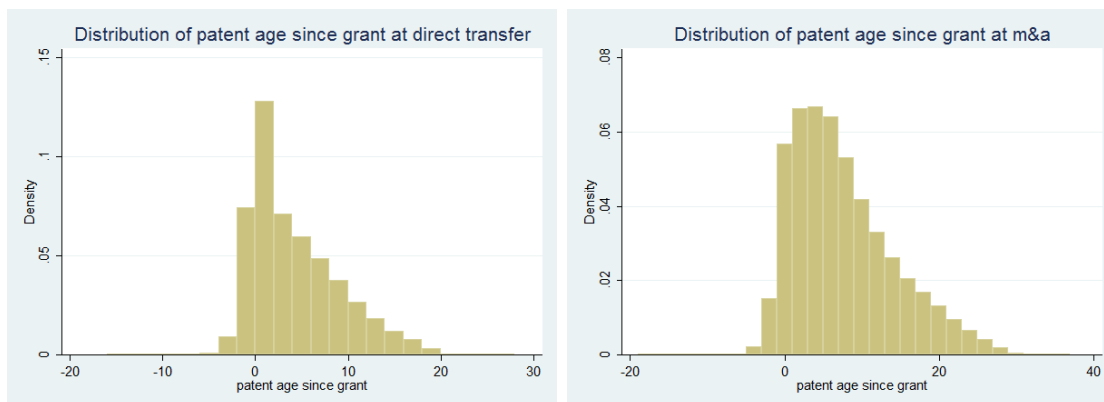
Figure 1.4: Patent age since grant. Up: Direct transfer. Down: M&a.

These results are consistent with the view that patents transacted with mergers and acquisitions are more important and have a more specific purpose. Transaction of patents transfers the tangible asset such as the production technology, as well as the intangible asset, such as the tacit knowledge. Transfer of the intangible asset has been studied by many researchers in the patent literature. However intangible assets embedded partially in the patent. If it were not, we would not observe a significant share of patent reallocation that takes place through mergers and acquisitions. For instance, there might be crucial information on replicating the patent or possibilities for further extensions and improvements. This information is not contained in the patent, which is a vague certificate or manual for a product or a production technology in the state as it is. From the contract theory point of view, this related tacit knowledge is not contractable. A Change of ownership or boundary of the firm is a natural tool to trade this non contractable resources.[4] This partially explains why we observe such a high rate of patents transferred with mergers and acquisitions. This is true particularly in areas like chemical, computers and electrical & electric areas, where the degree of complexity is high and it becomes easier to develop tacit knowledge about the technology or the innovation. It also takes longer to develop tacit knowledge, resulting in an average patent age that is higher with mergers and acquisitions. To support this hypothesis,

---

[4]Series works by Williamson (Wil10) and Grossman and Hart (1986)(GH86) have related discussions.

in the following section I show there is a strong(er) knowledge flow associated with mergers and acquisitions.

There is a caveat that in certain cases patents are not the purpose of mergers and acquisitions, while the intangible asset is. Nevertheless, studying information on the patents associated with mergers and acquisitions provides a scope to examine the information flows and transfer of tacit knowledge associated with mergers and acquisitions.

1.3   Analysis of the knowledge flow

Patent citations have been used by many authors to study knowledge flows geographically as in Jaffe, et al. (1993)(JTH93), and within sector as in Gittelman and Kogut (2003)(GK03), and also to study the information flow between firms, as in Singh and Agrawal (2011)(SA11). In subsection 1.3.1, I follow Singh and Agrawal (2011)(SA11) to examine the change of the internal citation rate of the patent pre- and post- transactions using a Difference-in-Differences approach. Change in the internal citation rate using different transaction methods was compared to each other and to the non-traded patents. A higher change in the internal citation rate suggests a higher utilization of the patent and is more associated with knowledge flows between firms. While citations measure the amount of information flow, I introduce the topic modeling to further study the content of the information flows in subsection 1.3.2. Results from both subsections support the hypothesis that there is a strong information flow associated with patent transactions, particularly with mergers and acquisitions.

1.3.1   Amount of knowledge flow

Citations and the citation rate are natural measures for patent usage as well as the information flow within or between firms. Less usage would suggest fewer internal citations, which is defined by the citations made within the same organization.

This leads to the following regression equation

$$Cites_{it} = f(X_{it}\beta + DD_{it} + \delta_i + I_{industry} + \epsilon_{it}), \tag{1.1}$$

where the dependent variable $cites_i t$ is the number of total internal citations received by patent i in year t (created by seller before the transaction, created by buyer after the transaction). The right hand side control variable X_{it} contains firm-level controls (previous patent size, previous patent size of the same class), patent age and whether the owner is a large entity. DD_{it} is a set of dummies, including: indicators of whether the patent is post-merger, post-direct-trade, or post any transactions and years after the type of the transaction. If there is an information flow associated with the patent transfer and as this information becomes more useful, we should observe higher coefficients for post-transaction period compared to the pre-transaction period. δ_i is the patent random effect. I also control the sector by having a two-digit sic sector dummy. $f(.)$ is chosen to be the negative binomial distribution for the count data to accommodate over dispersion in the data. This negative binomial kernel has been widely used in count-related data analysis. Cameron and Trivedi (2014)(CT14) provide a more detailed treatment. The regression results are presented in table (1.5).

Variables	$Cites_{it}$
year, industry dummies omitted	
age	0.0174***
	(0.0016)
age square	-0.0014***
	(0.0000758)
selling firm's patent size	$-6.88 \cdot 10^{-7}$***
	$(1.38 \cdot 10^{-7})$
buying firm's patent size in the same class	$1.66 \cdot 10^{-5}$***
	$(4.79 \cdot 10^{-6})$
big entity	0.0108***
	(0.0031)
transacted	-0.1319***
	(0.0326)
through m&a	0.1969***
	(0.0334)
post mergers year 1	0.1284***
	(0.0089)
post mergers year 2	0.1533***
	(0.0090)
post mergers year 3	0.1338***
	(0.0091)
post mergers year 4	0.1049***
	(0.0095)
post mergers year 5	0.0448***
	(0.0103)
post mergers year 6	0.0461***
	(0.0114)
through direct transfer	0.0794**
	(0.0355)
post direct transfer year 1	0.0539**
	(0.0172)
post direct transfer year 2	0.0638***
	(0.0183)
post direct transfer year 3	0.0989***
	(0.0187)
post direct transfer year 4	-0.0067
	(0.0201)
post direct transfer year 5	0.0195
	(0.0217)
post direct transfer year 6	-0.0096
	(0.0243)
constant	1.9544
	(0.0447)
Standard errors in parentheses	
*** p<0.01, ** p<0.05, * p<0.1	

Table 1.5: Citation rate changes.

After the transaction the citation rate drops significantly (captured by the negative coefficient before transacted). This is because the transaction pooled all different methods of

15

transactions. However if the patent is obtained through mergers and acquisitions, then the citation rate is significantly increased and the effect peaks around 2 years after the transaction. If the patent is obtained by direct transfer, then the citation rate barely increases (summing all the coefficients before transacted and direct transfer and years after direct transfer). Both the previous patent size and patents within the same class of the buyer show negative results since the more (similar) patents already owned by the buyer, the more substitutes the buyer has. The age effect shows a hump-shape on citation rates, which captures the diffusion curve. Chapter 2 of the dissertation is dedicated to modeling and estimating this diffusion curve.

To further examine the channel on how this intangible asset or the information is transferred, I examined the internal citations that were involved with the same original inventor. This could identify the channel where knowledge diffusion occurred. Zucker et al. (1998)(ZDB98) illustrated that star scientists from universities are important in shaping the US biotechnology firms. Singh and Agrawal (2011)(SA11) examined the effect of the inventor's flow in general. If the information is highly intangible, keeping the same inventors is a crucial part of the transaction to guarantee the successful transfer of the intangible asset. Those inventors will continue to stay after the mergers and acquisitions. This suggests the newly created citations will involve the same original inventors. This is called *acquire hire* in many industries.

Consider the following regression:

$$Selfcites_{it} = X_{it}\beta + DD_{it} + \delta_i + I_{industry} + \epsilon_{it}, \tag{1.2}$$

where the dependent variable *selfcites* is the total number of internal citations involved with the same inventor received by patent i in year t. DD_{it} is the set of dummies, including: indicators of whether the citation is post merger, post any transactions and also years post the type of the transaction and sizes. If there is *acquire higher* behavior, we should observe positive self-citations in the internal citations after the mergers.

Variable	$Selfcites_{it}$
age	-0.04552***
	(0.0005)
age square	0.0018***
	(0.0000322)
selling firm's patent size	$-2.53 \cdot 10^{-6}$***
	$(6.56 \cdot 10^{-8})$
buying firm's patent size in the same class	$1.19 \cdot 10^{-4}$***
	$(2.51 \cdot 10^{-6})$
big entity	-0.0136***
	(0.0014)
transacted	-0.0888***
	(0.0156)
through m&a	0.0488***
	(0.0160)
post mergers year 1	-0.1085***
	(0.0044)
post mergers year 2	-0.0880***
	(0.0044)
post mergers year 3	-0.0638***
	(0.0045)
post mergers year 4	-0.0506***
	(0.0047)
post mergers year 5	-0.0388***
	(0.0050)
post mergers year 6	-0.0390***
	(0.0055)
constant	0.5248***
	(0.0238)

Table 1.6: Self-citations.

Table 1.6 presents the regression results. After the m&a, the number of the self citations drops significantly. This is in line with some previous findings about the high inventors' turnover rate after the merger as in Seru (2014)(Ser14). However, this rate did not drop to zero immediately. There are inventors who are kept after the mergers and acquisitions and who continue with the R&D process. Meanwhile, the self-citation rate is also different from 1. This suggests there are researchers in the acquiring firm who join the research team and

work on the continuing R&D projects. The intangible assets or knowledge diffuse through this team work.

1.3.2   Content of the knowledge flow

In the previous subsection I showed there is an information flow associated with the transfer of the firms. In this subsection I study how this information flow is reflected in the change of the innovation direction. If the acquiring firms acquire the target firm to gain access to a specific technology or to extend their scope, they will gain new information outside their previous expertise. The newly acquired expertise will divert and broaden their research directions. The acquiring firm's patenting behavior reflects this change. Examining the change in the patenting direction identifies the existence of this information flow. The post-merger patenting direction should deviate from the pre-merger patents and be closer to the patents acquired in the mergers and acquisitions. I adopt the tool of topic modeling to quantify the patenting direction.

The topic modeling has been an important tool to understand the clustering of the text content. It has recently been used in empirical corporate finance, as in Hoberg and Phillips (2010) (HP10) in analyzing the SEC filings of public firms to quantify their product differentiation.

I use the non-negative matrix factorization method to cluster the content in patent title, abstract and descriptions. A detailed description of the algorithm can found in Xu et al. (2003)(XLG03). The non-negative matrix factorization has all weights being non-negative and thus avoids the difficulties in the interpretation of the negative weights of the articles in other clustering algorithms. The basic idea is the following: consider the fitting equation

$$Y = UV, \tag{1.3}$$

where Y is the text information in the file, U is an $m \times k$ matrix representing the k topics where each topic has m key words. V is a $k \times n$ matrix where n is the number of files in Y. V represents the share of each file in the k different topics. (V_{kn} becomes the coordinates of each file in the k-dimensional space.) V is constrained to be non-negative. The similarity

between any two files could be defined by the cosine metric between the two coordinates

$$\rho(i,j) = \frac{<V_{ki}, V_{kj}>}{|V_{ki}||V_{kj}|}. \tag{1.4}$$

Because the coordinates are positive, the cosine value will be in $[0, 1]$. The higher the cosine value, the higher the similarity $\rho(i,j)$. $\rho(i,j)$ will be in $[0, 1]$.

Introducing a text-based distance metric is better than the traditional patent metric based on similarities of the patent classes used Hall, Jaffe and Tratjenberg (1998) (HJT98) and Akcigit (2014)(ACG14). However many of the mergers and acquisitions in the high-tech fields happens between firms in the same industry or even in a similar area of research. The patents between the acquiring firms and target firms have significant overlap or belong to the same class. Topic modeling can break a class into subtopics and study the detailed R&D direction by its informational content. Secondly, traditional classification only shows whether a patent belongs a class or a subclass. Text-based distance metric shows how much a patent belongs to a topic and generates more variation in patent-level distances.

I focus on the Nano field (US class 977) because of its importance and high degree of tacit knowledge. To generate the topic clusters, I use the information from patent title, claim and abstract of the patents. There are 50 topics generated, and these are listed in Figure A.1-A.6. The weight matrix is then used to calculate the cosine similarity as described above.

Figure 1.5 plots the distribution of the pairwise similarities within a firm of two internally developed patents. The red solid line represents the firm without any external acquisitions of patents of any form. The blue dotted line represents the case for firms with m&as. The green dashed line represents the case for firms with external direct purchase only. In the last two cases, the m&a or direct external purchase of the patents happened between the two internally developed patents.
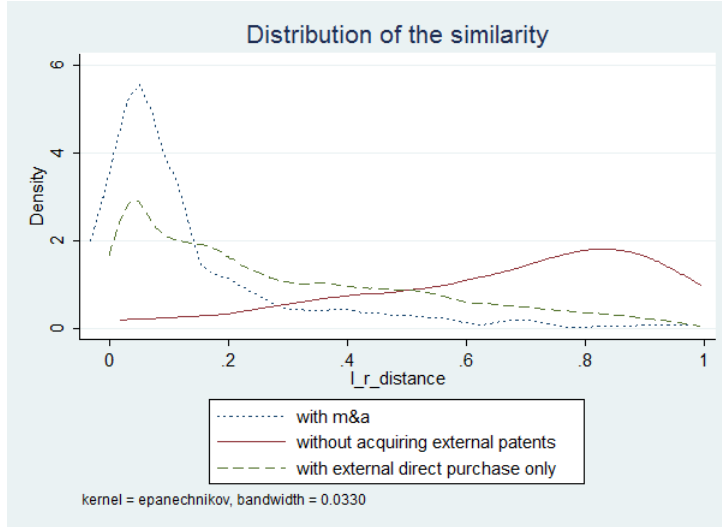
Figure 1.5: Overall similarity.

The firms without any external purchase show high persistence in their patenting direction. More mass is concentrated at the high similarity region. Firms associated with mergers and acquisitions have more distant patenting directions. However this does not show whether the new direction is related to the mergers and acquisitions behavior. I further break the similarity into the similarity between the internally developed patents before the mergers and acquisitions to the patents obtained through mergers and acquisitions, and the similarity between the internally developed patents post the mergers and acquisitions to the patents obtained through mergers and acquisitions.

Figure 1.6 plots the distribution of the similarities of the pre-merger internally developed patents to the patents acquired through mergers and acquisitions. The dashed line represents the similarities between the pre-merger internally developed patents and the post-merger internally developed patents. The dotted line represents the similarities between the post-merger internally developed patents and the patents acquired through mergers and acquisitions. The solid line represents the similarities between the pre-merger internally developed patents and the patents acquired through mergers and acquisitions. In all three cases, the similarity is low. This seems to be not supporting the hypothesis that the information flow associated with the mergers and acquisitions would bring the future patenting

20

direction closer to the patents acquired in the mergers and acquisitions.
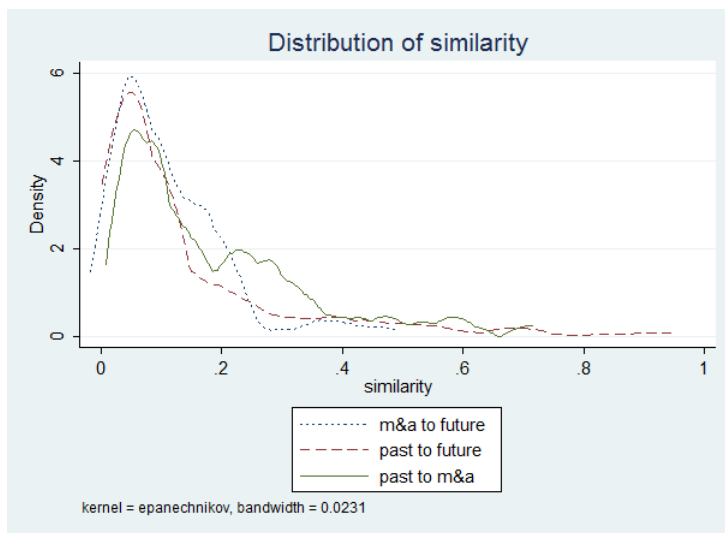
Figure 1.6: Similarity for mergers for all time periods.

However, knowledge diffusion takes time. Integrating the new information requires researchers at the acquiring firm to learn from the acquire hires and conduct experiments. This diffusion process normally takes roughly 3 to 5 years by examining the citation peaks. Taking this into consideration, Figure 1.7 plots the similarities with the internally developed patents at least 3 years post the merger.
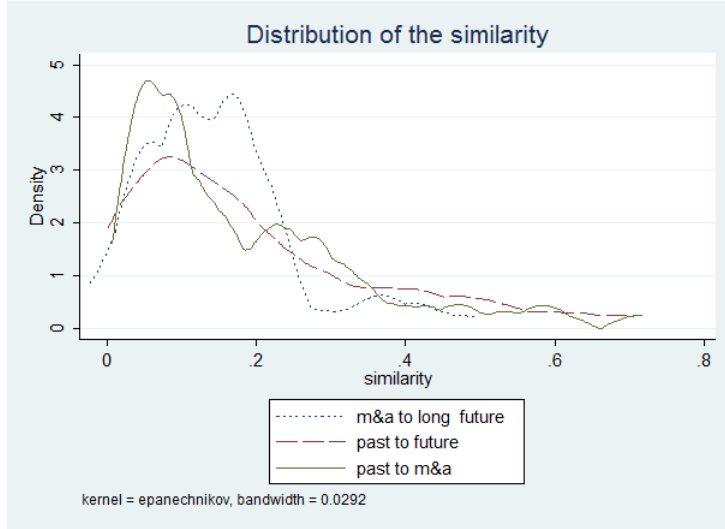
Figure 1.7: Similarity for mergers with long run effect.

Then we find that the long run similarity between post-merger internally developed patents and the patents acquired through mergers and acquisitions increases significantly. The similarity between post-merger patents and the pre-merger patents are still low. This confirms the role of mergers and acquisitions in changing the research direction and inducing information flow between firms.

One limitation of the current analysis is that I only show the existence of the information flow without showing the causality of this information flow. It is likely that acquiring firm decided to change their R&D direction before acquiring the target firm. This could be captured by scientific article citations and common coauthor-ships ahead of mergers and acquisitions could provide evidence, which renders future studies.

1.4   Conclusion

Patent reallocation is an important phenomenon. This paper summarizes some basic findings in which patents are transacted and how they are transacted. By examining the differences between different transaction methods, I show there is an information flow associated through mergers and acquisitions. Transacting firms is an alternative to inducing

information flow between firms besides patent transactions or citations. More formal economic modeling renders further consideration.

CHAPTER 2

Innovation or learning

2.1 Introduction

Innovation and learning are two important forces that drive economic growth. Patent policies might encourage innovation but impede learning. R&D subsidies might promote innovation at a high social cost if innovation is less important than learning. It is important to distinguish the relative influence of these two forces before introducing any targeted policy interventions to promote economic growth. Their relative contributions hinge on individual choices, which depend on the cost of innovation and the environment in which the learning takes place. Since knowledge flows from the person with more to the person with less, the option value of learning thus depends on one's current knowledge level and the entire distribution of others from whom he might learn from. This dependence will affect the innovation choices of the agents. On the other hand innovation choices will affect the evolution of this distribution. Identifying the cost structure and the learning environment become difficult since we have to take the dependence and feedback structure into account. With the recent seminal paper by Lucas and Moll (2014)(LM14) and developments in mean field game theory, these interactions can be traced explicitly. Using this framework we can separate innovation and learning and assess their relative contributions.

In our model, growth is generated by the accumulation of individuals' knowledge through learning or innovation. On the one hand, agents learn through interacting with other persons, which is modeled as random meetings at a constant rate. As each person meets others, knowledge flows from the person with more to the person with less. On the other hand, innovation is a process through which agents accumulate knowledge through costly research

and development activity. The state of the economy is fully summarized by the distribution of the knowledge. An agent's choice of R&D spending will be related to this distribution because the option value of meeting others affects the marginal payoff of innovation. The agents' R&D intensities and the rate at which they meet will determine the evolution of the knowledge distribution.

We first characterize the solution of the problem and then estimate the model by linking the learning process to the patent citation data. There are two equilibrium conditions: 1) the Bellman equation in which each agent chooses the R&D intensity with the knowledge distribution as given and 2) the law of motion equation, which describes the evolution of the knowledge distribution, given the meeting rate and the equilibrium R&D choices of each agent. We will focus on the balanced growth path.

The identification and estimation procedure consists of two steps. The key, and first step, is to study the inverse problem of the law of motion equation: given the distribution of knowledge, reconstruct the equilibrium R&D choices of the individual agents. An insight from Aït-Sahalia (1996)(Ait96) applies: given the observed distribution one can recover elements in the law of motion equation. Based on random walk approximation, the discrete Markov transition matrix can be replaced with the local first and second moments in the continuous time law of motion equation. Given the observed distribution, the law of motion equation pins down the relationship between these two moments. From knowledge of one, the other one can be reconstructed non-parametrically. If there is a monotonic relationship between this reconstructed moment and the equilibrium R&D choice, then the equilibrium R&D choice is identified. Using this method, we recover the equilibrium R&D choices from the distribution of knowledge, which is estimated from the patent citation data. Citations reflect the relative knowledge level embedded in the patent and therefore provide a source to estimate the distribution of the knowledge.

The second step involves an inverse optimal control problem: given the equilibrium R&D choices, reconstruct the parameters of cost functions in the Bellman equation. This second step has been studied in the context of dynamic discrete choice problems in the empirical

IO literature, although most of these studies estimated equilibrium choices directly from micro-level decision data.

Finally, we discuss how this two-step algorithm provides direct inference for a class of continuous time macroeconomic models such as the recent models in Lucas and Moll (2014)(LM14) and Luttmer (2007)(Lut07) and many other related continuous time (dynamic game) models.[1]

Our main results suggest the cost of innovation shows non-convexity in an agent's knowledge level. We find that the relative contribution of innovation to economic growth is more than 99 percent. Our result also explains the inverted-U shape relationship between the patent citations and the patent value documented recently in Abrams et al. (2014)(AAP14).

Related literature

My research contributes to two main research areas. In relation to the growth literature, we provide the first empirical strategy for a class of models studying the interactive effects in the technology diffusion. Our model is closely related to Lucas and Moll (2014)(LM14), Lutter (2013)(Lut13) and Jovanovic and Rob (1986)(JR86). In those models, technology diffuses through a random meeting process. Perla and Tonetti (2014)(PT14) had a similar framework to Lucas and Moll (2014)(LM14), but in their model learning only depends on the distribution in the economy and is independent of agent's knowledge level. We introduce an innovation choice to this class of model. This modeling environment has also been applied to the trade literature as in Alvarez et al. (2013)(ABR13) and Waugh et al. (2013)(WTP13). However there is no discussion of the empirical estimation. We provide the estimation strategy and it further allows for non-parametric identification.

Secondly, in relation to the empirical IO literature, we extend the standard two-step estimator in dynamic discrete choice models such as the seminal work in Bajari et al. (2007)(BBL07) or Pakes et al. (2007)(POB07) to continuous time. In these studies, the policy function is estimated from the micro-level data. By contrast we explore the implications from the macro density in the first step estimation. In this class of models we are

---

[1]Achdou et al. (2014)(ABL14) has a summary.

studying, since the number of the agents is large, the law of motion of the macro distribution of the states coincides with the law of motion of the probabilities of each agent's states. Then we can recover the individual choice from an inversion to the macro distribution. The second step is different from the conditional choice probability (CCP) estimators where an inversion from the choice probability to the value function based on the error structure is used to identify the value functions. In continuous time continuous states environment, derivative of the value function could be solved directly from a differential equation of recovered optimal choice. This two-step algorithm can be extended to a class of continuous time macroeconomic models or continuous time dynamic game models to study entry/exit, reallocation inefficiency, and other related topics. It also bridges a link between macroeconomic literature and the empirical IO literature through studying the dynamic choice problem. We will revisit this point in the conclusion.

Our study is also related to the study by Eeckhout and Jovanovic (2002)(EJ02) on the technology diffusion and the capital investment decision. They inferred the learning technology from the firm size distribution and compared results with the patent citation data in Caballero and Jaffe(1993)(CJ93). These studies assumed away the unobserved individual heterogeneity in the patent citation data. In our study, this heterogeneity is a means of identifying the knowledge distribution. In addition, by linking the patent citation data to the learning process we can also explain an inverted-U shape relationship between the patent citations and the patent value observed recently in Abrams et al.(2014)(AAP14). Apart from adding to the broad empirical literature on the estimation of the R&D cost, this paper also contributes to the literature on estimating the values in the patents. Among those studies, Pakes (1986)(Pak86) used the stopping time approach to infer the patent value from the maintenance fee events. Kogan et al. (2012)(KPS12) used a stock market valuation method. We explore a channel based on the dynamic change of the citations.

The paper proceeds as follows: Section 2 introduces the environment. In Section 3, we solve the equilibrium. Identification and estimation are addressed in Section 4. Results are discussed in Section 5. We conclude in Section 6.

2.2 Environment

The economy consists of a unit measure of infinitely lived agents. The population is defined by a cumulative distribution function:

$$F(x,t) = Pr\{s \leq x \text{ at date } t\}.$$

s, x is an index of knowledge level. $f(x,t)$ is the p.d.f function.

In each period the agent has a profit inflow of $\pi(x,t)$, which depends on the current level of knowledge x. He has limited amount of total effort. The agent chooses to spend e of the efforts on R&D with some disutility $C(e,t)$. The R&D effort accumulates the knowledge the agent has deterministically.

Conditional on his state is x at time t, agent maximizes the expected discounted flow of the profit stream

$$V(x,t) = E_t\{\int_t^\infty e^{-\rho(k-t)}[\pi(x_k, k) - C(e_k, k)]dk|x_t = x\},$$

where ρ is the discount factor.

The evolution of the distribution $F(x,t)$ was determined by two forces. Learning by random meeting: with probability α, he will meet another person x' in dt period and adopt the $\max(x, x')$. Innovation: an agent can increase his knowledge by an amount of $\mu(e(x), t)dt$ in the dt period by the R&D effort. To simplify the discussion, we allow the agent to directly choose the growth drift $\mu(x,t)$ with a transformed cost function $\Psi(\mu(.))$. $\mu(x,t)$ is less than or equal to some constraint $K(x,t)$ due to the limited amount of total effort.

The change in agent's knowledge between t and $t + dt$ could then be represented by

$$dx = \mu(x,t)dt + \Delta_t(x)dN_t.$$

The first term represents the drift chosen by the innovation effort. The second term N_t is a Poisson jump process of size $\Delta_t(x)$. It represents the effect by learning.

The value function could be rewritten as

$$\rho V(x,t) = \max_{\{\mu(x,t) \le K(x,t)\}} \pi(x) - \Psi(\mu(x,t)) + \partial_x V(x,t)\mu(x,t) + \partial_t V(x,t) \tag{2.1}$$

$$+\alpha \int_x [V(s,t) - V(x,t)]f(s,t)ds.$$

The first two term represent the flow profit and cost. The last term on the right reflects the option value of learning. Since agent can only learn from other agents with more advanced knowledge, the lower bound of the integral is x. α is the meeting rate.

As number of agents goes to infinity, the law of motion of $f(x,t)$ in the economy could be summarized through Kolmogorov forward equation[2] as

$$\frac{\partial f(x,t)}{\partial t} = -\alpha f(x,t) \int_x^\infty f(s)ds \bigg|_{out} + \alpha f(x,t) \int_0^x f(s)ds \bigg|_{in} \tag{2.2}$$

$$-\partial_x(\mu(x,t)f(x,t)) \bigg|_{R\&D}.$$

This characterization of the law of motion is a key feature in Mean field games related models. In these models, since the agents are symmetric and the probability of each individual states coincides with the macro distribution due to law of large number. Then we can only trace the evolution of the macro distribution.

Its $c.d.f$ form is

$$\frac{\partial F(x,t)}{\partial t} = -\alpha F(x,t)(1 - F(x,t)) \bigg|_{learning} - \mu(x,t)f(x,t) \bigg|_{innovation}. \tag{2.3}$$

We will use this $c.d.f$ representation of the law of motion equation through out the paper. The left hand side is the total change in the population below x at time t. At any time t, there are two ways to exit the pool of the agents below x. The first is meeting with another agent that is higher than x. This happens with probability $\alpha(1 - F(x,t))$ and explains the first term on the right. The second term is exit by innovation. In infinitesimal time, only the marginal agents $\mu(x,t)f(x,t)$ can exit.

---

[2]Details are in Appendix A.2.2.

2.3   Solution and equilibrium

An equilibrium, given the initial distribution $f(x,0)$, is a triple $(f(), \mu(), V())$ of functions on R^2_+ such that (i) given $\mu(x,t)$, $f(x,t)$ satisfies the law of motion equation (2.2) for all (x,t); (ii) given $f(x,t)$, $V(x,t)$ satisfies the Bellman equation; and (iii) $V(x,t)$ attains its maximum at $\mu(x,t)$.

A balanced growth path is a rate λ and functions $(\tilde{F}(), v(), \bar{\mu}(), \bar{\pi}(), \bar{\Psi}())$ on R such that

$$
\begin{aligned}
F(x,t) &= \tilde{F}(xe^{-\lambda t}), \\
V(x,t) &= e^{\lambda t} v(xe^{-\lambda t}), \\
\mu(x,t) &= e^{\lambda t} \bar{\mu}(xe^{-\lambda t}), \\
\pi(x,t) &= e^{\lambda t} \bar{\pi}(xe^{-\lambda t}), \\
\Psi(\mu(x,t)) &= e^{\lambda t} \bar{\Psi}(\mu(xe^{-\lambda t})),
\end{aligned}
$$

and for all (x,t), $(f(), \mu(), V())$ is an equilibrium given the initial condition: $F(x,0) = \tilde{F}(x)$.

Define $z \equiv xe^{-\lambda t}$ and relabel

$$
\begin{aligned}
\tilde{F}(xe^{-\lambda t}) &= \tilde{F}(z), \\
v(xe^{-\lambda t}) &= v(z), \\
\bar{\pi}(xe^{-\lambda t}) &= \bar{\pi}(z), \\
\bar{\mu}(xe^{-\lambda t}) &= \bar{\mu}(z), \\
\Psi(\mu(xe^{-\lambda t})) &= \bar{\Psi}(\bar{\mu}(z)).
\end{aligned}
$$

$v(z)$ is the rescaled value function. $\tilde{F}(z)$ is the cumulative distribution of the knowledge on the balanced growth path. $\bar{\mu}(z)$ is the rescaled optimal choice. $\bar{\Psi}(\bar{\mu}(z))$ is the rescaled cost function. Now the Bellman equation (2.1) becomes

$$(\rho - \lambda)v(z) = \max_{\bar{\mu}(z) \leq \tilde{K}(z)} \bar{\pi}(z) - \bar{\Psi}(\bar{\mu}(z)) + (\bar{\mu}(z) - \lambda z)v'(z) + \alpha \int_z [v(z') - v(z)]\tilde{f}(z')dz'. \quad (2.4)$$

The law of motion equation (2.2) becomes a first-order nonlinear ordinary differential equation of $\tilde{F}(z)$,

$$\alpha \tilde{F}(z)(1 - \tilde{F}(z)) = (\lambda z - \bar{\mu}(z))\tilde{f}(z). \quad (2.5)$$

30

The left-hand side of equation (2.5) is the outflow of the agents below level z through learning. The right-hand side is the inflow of the agents that fall back to levels below z. In equilibrium, the two are balanced.[3] In the model we will impose the upper limit of the R&D effort $\bar{K}(z) = \lambda z$ and $\bar{\Psi}(\bar{K}(z)) = \infty$. This is in accordance with the non negativity of the right hand side of equation (2.5).

Proposition 1. *If $\widetilde{\mu(z)} \equiv \lambda z - \bar{\mu}(z)$ in which $\widetilde{\mu(z)} \in [0, \lambda z]$, satisfies the condition: $\lim_{s \to \infty} \int_{\bar{z}}^{s} \frac{1}{\widetilde{\mu(z)}} = +\infty$, e.g. $\widetilde{\mu(z)} \sim \mathcal{O}(z^d)$ and $d \leq 1$, then the solution to equation (2.5) is $\tilde{F}(z) = \frac{1}{1 + e^{M(z)}}$ where $M(z) = -\int_{\bar{z}}^{z} \frac{\alpha}{\widetilde{\mu(s)}} ds - lnC$ and $\tilde{F}(\bar{z}) = \frac{C}{1+C}$.*

Proof: See Appendix A.2.3

To guarantee the existence of balanced growth rate λ, we provide a sufficient condition:

Lemma 1. *If the initial distribution satisfies $\lim_{z \to \infty} \frac{z(1-\tilde{F}(z))'}{1-\tilde{F}(z)} = -\omega \leq 0$ and the equilibrium satisfies $\lim_{z \to \infty} \frac{\bar{\mu}(z)}{z} = \eta$, then there exists a balanced growth path rate $\lambda = \frac{\alpha}{\omega} + \eta$.*

Proof: See Appendix A.2.4

If $\lim_{z \to \infty} \frac{z(1-\tilde{F}(z))'}{1-\tilde{F}(z)} = -\omega \leq 0$, then by the Karamata theorem ((Fel71)), $1 - \tilde{F}(z)$ has a regular varying component $z^\omega l(z)$, where $l(z)$ is a slowly varying function. This means $\tilde{F}(z)$ has a fat tail with tail index ω.

---

[3] Several variants of this equation have been studied in Lucas and Moll (2014)(LM14), Alvarez et al. (2008)(ABL08) and Luttmer (2013)(Lut13).
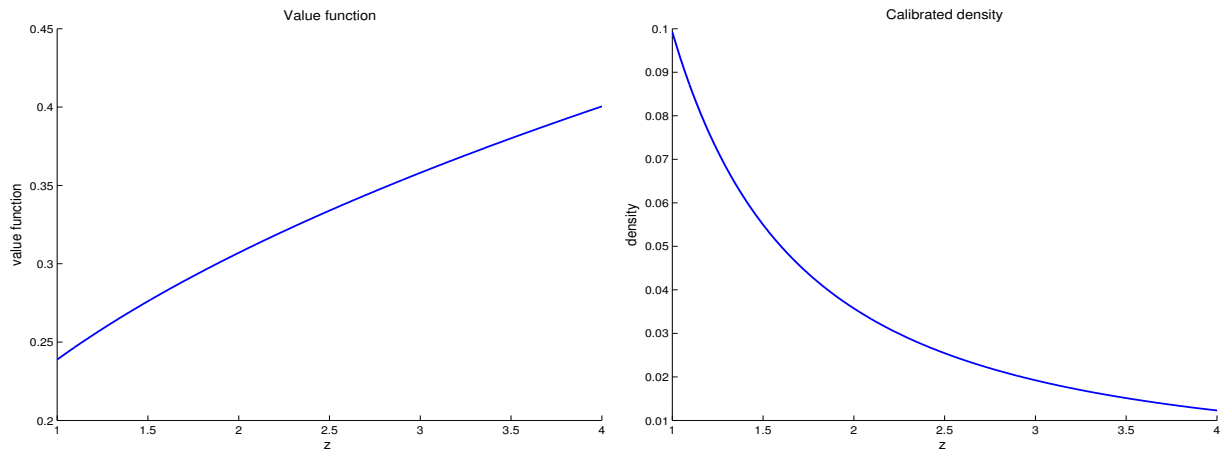
Figure 2.1: Calibrated value function.

The equilibrium density is similar to a Pareto distribution. Growth rate $\lambda = 0.02$, $\alpha = 0.3$, tail

index $\omega = \frac{\lambda}{\alpha}$, $C = 0.001$
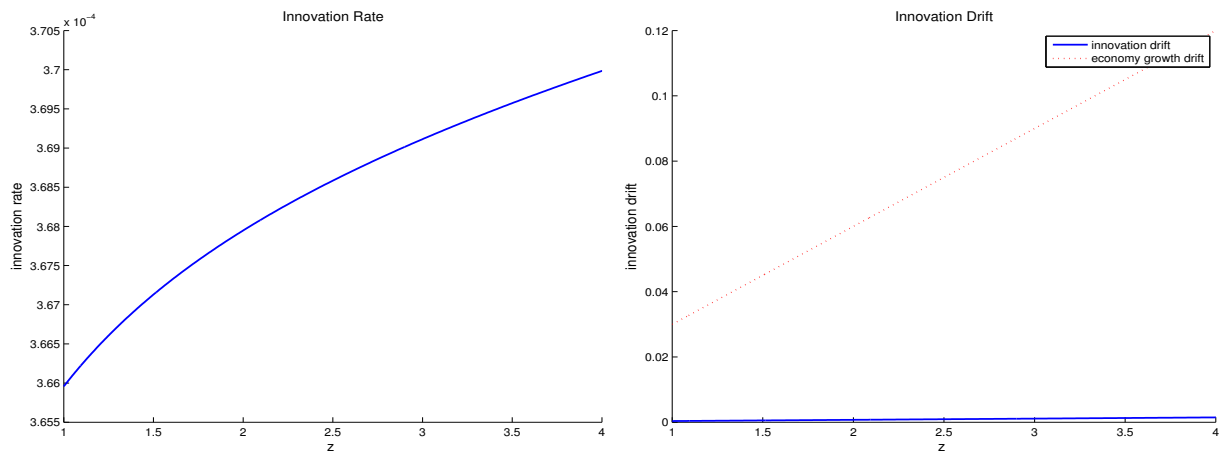
Figure 2.2: Calibrated innovation choice.

The equilibrium choice of the drift. Innovation rate is defined as $\frac{\bar{\mu}(z)}{z}$.

2.4   Empirical strategy

The essential goal is to recover the cost function $\bar{\Psi}(.)$.

2.4.1 Two-step algorithm and identification

In the first step, we recover the equilibrium R&D choices $\bar{\mu}(z)$ from the balanced growth path distribution of the knowledge $\tilde{F}(z)$. If $\tilde{F}(z)$ was observed and the growth rate λ and the meeting rate α were known, then the optimal R&D choice $\bar{\mu}(z)$ could be identified and estimated from equation (2.5) directly by

$$\bar{\mu}(z) = \alpha \frac{\tilde{F}(z)(1 - \tilde{F}(z))}{\tilde{f}(z)} - \lambda z.$$

The insight is similar to the result of Aït-Sahalia (1996)(Ait96). Given the observed distribution, one can reconstruct the element in the continuous time law of motion equation. Here we abstract from Brownian motion and the inversion is analytical. In the appendix, we include more examples of inversion to different types of law of motion equations. [4]

In a second step, with the empirical knowledge of $\bar{\mu}(z)$ and $\tilde{F}(z)$, $v(z)$ could be derived from (2.4) up to a specification of $\bar{\pi}(z)$ through an inverse optimal control problem. Bellman (1970)(Bel70) and Casti (1980)(Cas80) provided the general solution to this problem in a non-stochastic case. Chang (1988)(Cha88) studied the stochastic case in a growth application. If we take the derivative of the Bellman equation (2.4) with respect to z one more time and substitute in the envelope condition,

$$(\rho - \lambda)v'(z) = \bar{\pi}'(z) + (\bar{\mu}(z) - \lambda z)v''(z) - \lambda v'(z) - \alpha(1 - \tilde{F}(z))v'(z).$$

By rearranging the terms and substituting in the first-order condition we have:

$$(\rho + \alpha(1 - \tilde{F}(z)))v'(z) \quad = \bar{\pi}'(z) + (\bar{\mu}(z) - \lambda z)v''(z), \tag{2.6}$$

$$(\rho + \alpha(1 - \tilde{F}(z)))\bar{\Psi}'(\bar{\mu}(z)) \quad = \bar{\pi}'(z) + (\bar{\mu}(z) - \lambda z)\bar{\Psi}''(\bar{\mu}(z))\bar{\mu}'(z). \tag{2.7}$$

We provide the solution to equation (2.6) and (2.7) by solving these ordinary differential equations. Denote $L(s) = \frac{\rho + \alpha(1 - \tilde{F}(s))}{\lambda s - \bar{\mu}(s)}$, then (2.6) becomes

$$L(z)v'(z) + v''(z) = \frac{\bar{\pi}'(z)}{\lambda z - \bar{\mu}(z)}, \tag{2.8}$$

---

[4]The idea of reconstructing coefficients in differential equations applies to many inverse problems.

Its solution is

$$v'(z) = e^{-\int^z L(s)ds} \int^z e^{\int^\tau L(s)ds} \frac{\bar{\pi}'(\tau)}{\lambda\tau - \bar{\mu}(\tau)} d\tau + C_1.$$

By the first-order condition $\bar{\Psi}'(\bar{\mu}(z)) = v'(z)$,

$$\bar{\Psi}(\bar{\mu}(z)) = \int^z v'(s)\bar{\mu}'(s)ds + C_2.$$

If $\bar{\mu}(z)$ and $\tilde{F}(z), \lambda, \alpha$ are recovered $v(z)$, and $\bar{\Psi}(\bar{\mu}(z))$ are recovered up to normalizations C_1, C_2 and the specification of $\pi(z)$. In Appendix A.2.12, we provide a numerical method based on the finite difference method.

Alternatively, one can proceed and use forward simulations to simulate $v(z)$ given some parametric assumptions θ of the cost function $\bar{\Psi}(\bar{\mu}(z); \theta)$. Then one finds optimal parameters through the minimum distance estimator, θ_0, which minimizes the fitting difference in either (2.6) or (2.7). This resembles the strategy of the two-step estimator in Bajari et al. (2007)(BBL07)(B.B.L.).

This algorithm extends the two-step estimator in dynamic discrete choice models to the continuous state and continuous time environment. In a standard two-step estimator, the first step will recover the policy function from the micro-level decision data. In this continuous time setup, the moments need to be recovered are greatly reduced because of the random walk structure. Secondly, since the number of the agents is large enough, the law of motion equation of the macro distribution of the states coincides with the state probability of each agent. Then, an inversion to the macro distribution is the same as to the individual state probability.

2.4.2   Identification of $\tilde{F}(z)$

There are relatively few studies trying to directly estimate $\tilde{F}(z)$, the distribution of knowledge within the diffusion framework, except for Eeckhout and Jovanovic (2002)(EJ02). They link $\tilde{F}(z)$ to the empirical distribution of book values of the firms by assuming firm size is proportional to the knowledge embedded in them. They then compared the inferred learn-

ing technology with the aggregate patent citation data in Caballero and Jaffe (1993)(CJ93). This is similar to inferring the meeting rate α from the empirical knowledge of $\tilde{F}(z)$ in our environment (equation (2.5)). Whether knowledge is proportional to size depends on the assumptions on $\bar{\pi}(z)$. We will infer $\tilde{F}(z), \alpha, \lambda$ and the constant C that regulate the boundary mass directly from the patent citations data without this proportional assumption.

2.4.2.1 Assumptions

To estimate the unobserved $\tilde{F}(z)$, we link it to the patent citations by the following assumptions:

Assumption 1. *When agents learn from others' knowledge, they make citations to the earlier knowledge from their original level. For instance, when agent x meets agent x', where $x < x'$, he will adopt x' and cite all the patents between x and x'.*
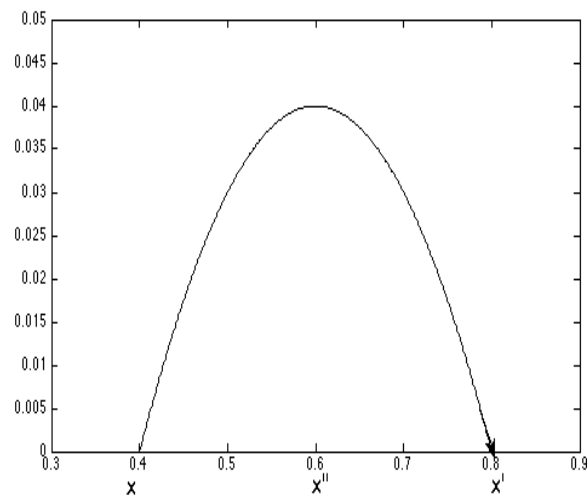

Figure 2.3: Citation creation process.

When x meets x', x creates a citation to patents between x and x' (not to ones below x).

Assumption 2. *The citation counts follow a Poisson process.*

According to guidelines from the United States Patent and Trademark Office (USPTO), a patent is only granted for sufficient improvements. Citations serve as a record of the

knowledge levels it went through. This suggests the citations received by a patent with level z_i followed a Poisson process with a rate proportional to $\alpha \tilde{F}(z_i e^{-\lambda t})(1 - \tilde{F}(z_i e^{-\lambda t}))$. The reason is that in each period t, there are $\tilde{F}(z_i e^{-\lambda t})$ agents below z_i. With probability $\alpha(1 - \tilde{F}(z_i e^{-\lambda t}))$, they surpass level z_i by meeting with agents who have better knowledge than z_i. The product $\alpha \tilde{F}(z_i e^{-\lambda t})(1 - \tilde{F}(z_i e^{-\lambda t})) f(z_i, t)$ is the total citations created for the level z_i patents. There are $f(z_i, t)$ patents of technology z_i. The average citations received is total citations created divided by total citations at that level.[5]

Assumption 3. *There is no correlation between z_i and other observed covariates.*

In each period, the citation rate of a patent is determined soley by its relative knowledge level in the economy. This is assumed for technical reasons to show identification. A similar assumption has been made in Gurmu et al. (1998)(GRS98).

Assumption 4. *Assume $C > 0$.*

Assumption 5. *The initial distribution of the knowledge in the patents coincides with the initial distribution of the knowledge in the economy.*

Assumption 6. *Patents are static. In the following period, relative knowledge level in a patent with level z_i becomes $z_i e^{-\lambda(t+1)}$ deterministically.* [6]

We make this assumption because a patent reflects the knowledge an agent has at time t. Agents evolve by learning or innovation, but the knowledge in that particular patent stays the same in the absolute level.

2.4.3 Identification steps

Through these assumptions, patent quality is reflected in the time-series behavior of its citations. The cross-sectional data of these different qualities will identify the distribution

---

[5] A similar assumption has been made in Griliches (1957)(Gri57)'s earlier studies and in Bertran (2006)(B06), where he assumed all the past patents would be studied and cited by an agent through a binomial process.

[6] *In Appendix A.2.5, we have a slightly different formulation.*

of the knowledge. These assumptions suggest we might gradually observe that some patents are with upward slopping shape, some will be hump-shaped and others are with downward slopping shape. The longer the upward slopping period is, the higher the quality. This is due to the difficulty of learning and making contribution to the advanced knowledge. We then can rank the patents accordingly and get the distribution $\tilde{F}(z)$.

Proposition 2. *Meeting rate α, lower bound constant C, rate of obsolescence λ, c.d.f $\tilde{F}(z)$ are identified under assumptions 1-6.*

To identify $\tilde{F}(z)$, α, λ, C, we have two main steps: i) the identification of the Poisson rates in the model; ii) the identification of $\tilde{F}(z)$ from the assumption of the function form of the Poisson rate (single peaking that holds overall for the empirical observation). We show the identification strategy by identifying: α, $\tilde{F}(0)$, the Poisson rate $m_{t,i}$, an interim object R_{it}, and $\tilde{F}(z)$ and λ.

2.4.3.1 Identifying meeting rate α and C $\quad (\tilde{F}(\bar{z}) = \frac{C}{1+C})$

Lemma 2. *α and C are identified.*

Under assumption 2, data follows Poisson process. The empirical characteristic function $\phi_H(s)$ for data in period 0 satisfies

$$\phi_H(s) = \int_{\bar{z}}^{\infty} \phi_k(\alpha \tilde{F}(z)(1 - \tilde{F}(z)), s) d\tilde{F}(z),$$

where $\phi_k(\alpha \tilde{F}(z)(1 - \tilde{F}(z)), s)$ is the characteristic function for poisson random variable with rate $\alpha \tilde{F}(z)(1 - \tilde{F}(z))$. It takes the form $\phi_k(\alpha \tilde{F}(z)(1 - \tilde{F}(z)), s) = e^{(e^{is}-1)\alpha \tilde{F}(z)(1-\tilde{F}(z))}$. Then we have

$$
\begin{aligned}
\phi_H(s) &= \int_{\bar{z}}^{\infty} \phi_k(\alpha \tilde{F}(z)(1 - \tilde{F}(z)), s) d\tilde{F}(z) \\
&= \int_{\bar{z}}^{\infty} e^{(e^{is}-1)\alpha \tilde{F}(z)(1-\tilde{F}(z))} d\tilde{F}(z) \\
&= \int_{\frac{C}{1+C}}^{1} e^{(e^{is}-1)\alpha w(1-w)} dw \\
&= \frac{\sqrt{\pi} e^{\frac{\alpha(e^{is}-1)}{4}} (erf(\frac{1}{2}\sqrt{\alpha}\sqrt{e^{is}-1})) - erf(\frac{1}{2}(2\frac{C}{1+C}-1)\sqrt{\alpha}\sqrt{e^{is}-1}))}{2\sqrt{\alpha}\sqrt{e^{is}-1}}
\end{aligned}
$$

37

The third line changes the variable from $\tilde{F}(z)$ to w. The lower bound of the integrand is $\tilde{F}(\bar{z}) = \frac{C}{1+C}$ and the upper bound is $\tilde{F}(\infty) = 1$. $erf(.)$ is the error function where $erf(x) = \frac{2}{\sqrt{\pi}} \int_0^x e^{-\frac{l^2}{2}} dl$. On the right hand side there are two unknowns: the meeting rate α and the lower bound $\frac{C}{1+C}$. This is an over identified system.

2.4.3.2 Identifying the Poisson rates $m_{t,i}$ and an interim object R_{it}

The possible Poisson rate $\{m\}$ and its probability $\Gamma(dm)$ in the economy are identified. The proof was adopted from Rao (1992)(Rao92) p214 and is listed in Appendix A.2.6. The procedure is similar to the steps of identifying α showed earlier but replacing $\alpha\tilde{F}(z)(1 - \tilde{F}(z))$ with m and $\tilde{F}(z)$ with $\Gamma(dm)$. Then the mixture of the characteristic function could be transformed to the Laplace transform of $\Gamma(dm)$. The identification follows from the uniqueness of the Laplace transformation.

The joint Poisson rates $m_0, m_1, ..m_T$ and the joint density $\Gamma(m_0, m_1, ..m_T)$ and the conditional probabilities are identified following Teicher (1967)(Tei67), which says if the marginals of the mixture are identified separately then the joint density of the mixture is identified. The details are listed in Appendix A.2.7. It is worth mentioning, each m_t is a vector of the possible rates m_{t_j} in period t.

Define a path i in $m_0, m_1, ..m_T$ by a $T+1$ tuple of $(m_{0,i}, m_{1,i}, ..m_{T,i})$ where $m_{t,i}$ is a rate picked from m_t. By assumption 2 for fixed z_i, the poisson rate function is

$$\alpha\tilde{F}(z_i e^{-\lambda t})(1 - \tilde{F}(z_i e^{-\lambda t})) = m_{t,i}. \tag{2.9}$$

This rate function is symmetric and single-peaked in $\tilde{F}(z_i e^{-\lambda t})$. On a path i varying t, $m_{t,i}$ can be one of the three possible shapes: upward sloping, a hump shape or downward sloping. Along the path, upward trends are associated with $\tilde{F}(z_i e^{-\lambda t}) > \frac{1}{2}$ and downward trends are associated with $\tilde{F}(z_i e^{-\lambda t}) < \frac{1}{2}$. These different shapes allow us to separate $\tilde{F}(z_i e^{-\lambda t})$ from $m_{t,i}$.

Define $R_{it}^+ = \frac{1+\sqrt{1-4\frac{m_{t,i}}{\alpha}}}{2}$ and $R_{it}^- = \frac{1-\sqrt{1-4\frac{m_{t,i}}{\alpha}}}{2}$. These are the two roots to the equation (2.9) for $\tilde{F}(.)$.

If $m_{t,i}$ is upward sloping, i.e. $\Gamma(m_{t+1,i} > m_{t,i}|m_{t,i}) > 0$ for all t, then $R_{it} = R_{it}^+$.

If $m_{t,i}$ is downward sloping, i.e. $\Gamma(m_{t+1,i} < m_{t,i}|m_{t,i}) > 0$ for all t, then $R_{it} = R_{it}^-$.

If there exists a turning point n such that before the turning point, $m_{t,i}$ increases with t and after the turning point $m_{t,i}$ decreases with t, then $R_{it} = R_{it}^+$ for t before n and $R_{it} = R_{it}^-$ for t after n .

Then we identified $\tilde{F}(z_i e^{-\lambda t}) = R_{it}$.[7]
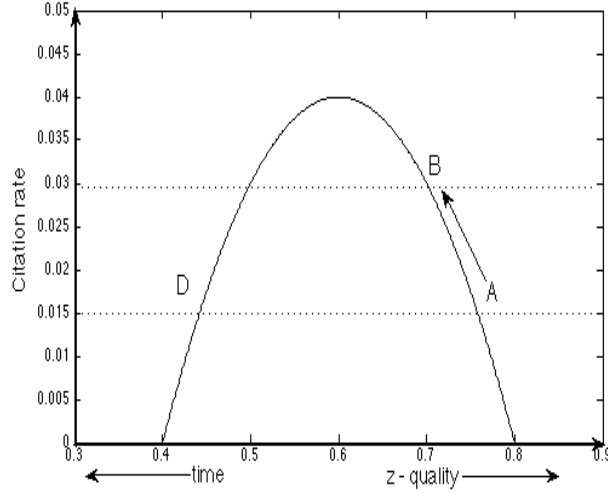
Figure 2.4: Identification of R_{it}.
If we observe $A \rightarrow B \rightarrow D$, then we can identify $\tilde{F}(z_i e^{-\lambda t})$ correctly along the path by finding

$$R_{it}.$$

2.4.3.3 Identifying $\tilde{F}(z)$ and rate os obsolescence λ

Lemma 3. *If there is some identified interim object R_{it} such that $R_{it} = \tilde{F}(z_i e^{-\lambda t})$ and α is known, then $\tilde{F}(z)$, λ are identified.*

In period 0, $\tilde{F}(z_i) = R_{i0}$. The density of R_{i0} coincides with $\tilde{f}(z)$ and $\tilde{F}(z)$ is identified.

For $t > 0$ we use the inverse of $\tilde{F}$ and take log

$$log(z_i) - \lambda t = log(\tilde{F}^{-1}(R_{it})) \qquad (2.10)$$

[7]We assume dt is small enough in the data.

λ is identified.

Essentially this is a nonseparable mixture model and a classic strategy could be found in Matzkin (2003)(Mat03). The intuition is that with an adequate choice of normalization, for some fixed $\hat{t}$ the variation in $R_{.|\hat{t}}$ will be the variation of z_i, and then both the distribution of z_i and $\tilde{F}(.)$ are identified.

This completes the identification steps.

2.4.4   Construction of the likelihood in the first step

For each period j, if the observed received citation for patent i is y_{ij} and its knowledge level is z_i, then this probability is

$$P_{ij}(y_{ij}|z_i) = e^{\alpha(\tilde{F}(z_i e^{-\lambda j}))(1-\tilde{F}((z_i e^{-\lambda j}))} \frac{[\alpha(\tilde{F}(z_i e^{-\lambda j}))(1-\tilde{F}((z_i e^{-\lambda j}))]^{y_{ij}}}{y_{ij}!}$$

. For the entire history of the T observation periods, we get $\Pi_{j=0}^{j=T} P_{ij}(y_{ij}|z_i)$. However, we don't observe z_i, this knowledge level is the object we try to infer. By our assumption (5), the distribution of z is the same as $\tilde{f}(z)$. Then the probability of the observations for i becomes

$$P(y_i) = \int \Pi_{j=0}^{j=T} P_{ij}(y_{ij}|z)\tilde{f}(z)dz, \tag{2.11}$$

and the log likelihood function is

$$P(y) = \sum_i^N log(P(y_{ij})). \tag{2.12}$$

We find $\hat{\tilde{\mu}}(z), \hat{\alpha}, \hat{C}$ and $\hat{\lambda}$ to maximize the log likelihood function (2.12). We use method of sieves (*hermite polynomials*) to approximate $\frac{1}{\tilde{\mu}(.)} \equiv [\Sigma_{k=0}^\infty \beta_k h_k(z)]^2$: $\tilde{F}(z) = \frac{1}{1+\frac{1}{C}e^{\alpha \int^z [\Sigma_{k=0}^\infty \beta_k h_k(s)]^2 ds}}$ and use the information criterion to determine the order of k. The use of orthogonal basis polynomials as sieves was introduced in Gallant and Nychka (1987)(GN87) as a sieve estimator [8]. Under some mild conditions on the local boundedness of the gradient function of

---

[8]Chen (2007)(Che07) summarizes the major methods and applications.

the likelihood, the likelihood function and the score function are continuous in the estimators. Then the maximum likelihood estimator is consistent. We use delta methods to give confidence intervals. Chen et al.(2014)(CLS14) and Chen and Liao (2014)(CL14) provide the analysis for the asymptotics of the plug-in sieve estimators.

Cmeraon and Johansson (2004)(CJ04) and Gurmu et al. (1998)(GRS98) applied a similar technique in a Poisson mixture setup. In those flexible models, they used a candidate density and adjusted it with quadratic polynomial basis functions to semiparametrically estimate the Poisson mixture. In our setup, the density form is parametric up to the knowledge of $\frac{1}{\widehat{\mu(z)}}$.

2.4.5 The second step estimation

We plug the estimates of $\hat{\tilde{\mu}}(z)$ into the second step by equations (2.6) and (2.7). We use the delta method to calculate the confidence interval.

If there are no explicit solutions to the inverse optimal control equation, we can express the derivative of the value function by sieve estimates. Then the second step and the first step could be jointly estimated through Generalized Method of the Moments with two moment conditions: the inverse optimal control equation (2.6) and the law of motion equation (2.5). The sieve parameters of the equilibrium choice will be treated as first stage parameters and the value function parameters will be the second stage parameters (see Ackerberg et al. (2012)(ACH12)). The estimator has asymptotic normality.

2.5 Results

2.5.1 Summary of the data

In this section, the data is briefly introduced. The main data source used is the NBER Patent Data Project (PDP) dataset. It provides the patent data and the patent citation data from 1976 to 2006. We use the citations received (forward citations) as the dependent variable y_{it}. There were approximately three million patents issued during this time period. We estimate the model by fixing a year cohort (year 1985). There were $67,405$ granted

patents that received citations in this period. They are classified into six categories.

Category	Total patents	Avg citations	std.
Chemical (excluding drugs)	13,502	10.32	1.1888
Computers & Communications	6,954	17.92	1.8194
Drugs & Medical	5,334	19.50	2.6461
Electrical & Electronics	11,868	11.68	1.3085
Mechanical	15,007	9.17	0.9866
Others	14,740	10.33	1.1216

Table 2.1: Patents data in summary.
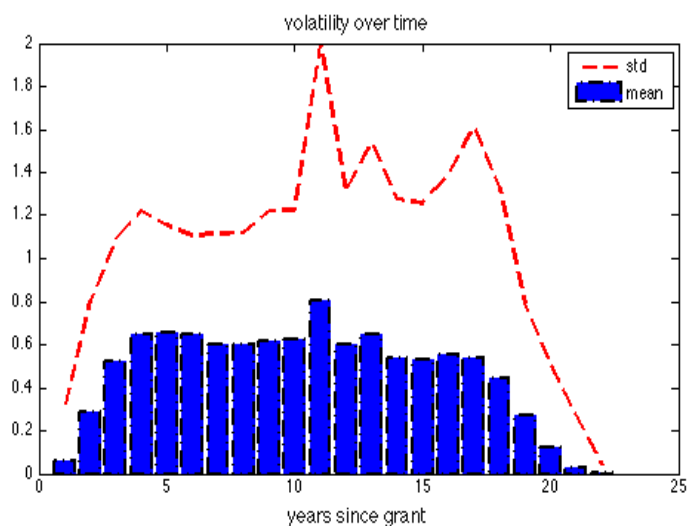
Figure 2.5: Citations received over time.

There is a significant amount of heterogeneity in citations received across patents. Table (2.1) summarizes the average total citations received and the standard deviation within each category. Patent citations in each category show a 5-10 % variation.

We focus on the chemical patents, which is the largest category. In Figure 2.5, we plot the average citation of the chemical patents over time, which is represented by the blue bar

42

charts. The red dashed line is the standard deviation of the citations. The average citations received show a hump-shaped change over time, which motivates the diffusion curve type regression in Caballero and Jaffe (1993)(CJ93) at an aggregate level. The time pattern of the standard deviations among citations received also shows a hump shape over time. This pattern suggests that the differences in the levels of knowledge in patents are reflected over time. Thus the panel data provides a source to estimate the distribution of knowledge.

2.5.2 Results

Parameters	Value	Standard deviation	Huber robust standard deviation
α	0.3954	0.0054	0.0043
C	0.9007	0.1516	0.5451
λ	0.1286	0.0084	0.0043

Table 2.2: Estimation Results.

Table (2.2) summarizes the major estimation results for the meeting intensity α, constant C, and discount rate λ. A smaller value of C means the degenerate mass at the lower bound is smaller. The estimated rate of obsolescence λ is about 12% per year. This is within the range of findings from Caballero and Jaffe (1993)(CJ93), which is around 8% to 16% per year. Similar to their framework, creative destruction in our model is related to the profit size. When the profit margin is larger, the incentive to innovate is stronger and the speed of obsolescence is faster. In contrast to their framework in which learning only affects the innovation technology, learning in our model contributes to the creative destruction directly in our environment. A stronger learning process increases the replacement effect. A second difference from their framework is that we study the heterogeneity of individual knowledge levels while they studied the aggregate pattern of the data. Some individual heterogeneity might be attributed to the rate of obsolescence in the macro data in their environment. This estimated rate of obsolescence is higher than the result of 2% from the Lucas and Moll (2014)(LM14) calibration exercise, in which they link the rate to the GDP growth rate of

43

major OECD countries (including the U.S.).

95% confidence inverval for estimated density
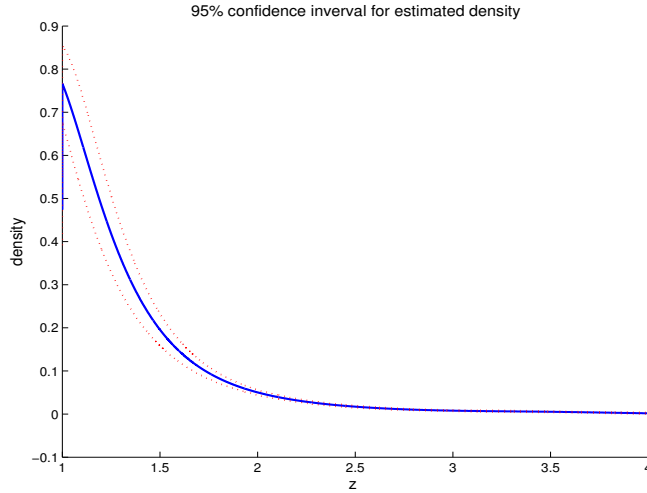
Estimated density $\tilde{f}(z)$ is illustrated by the solid blue line and the 95% confidence interval by the red dotted line.

Figure 2.6: Estimated density.

Percentile	Value
5	lower bound
25	lower bound
50	1.0339
75	1.4788
mean	1.7156
95	2.64

Table 2.3: Dispersion of z.

The distribution of knowledge is similar to a Pareto distribution. There is over dispersion and the mean/median ratio for the knowledge level is 1.66. This dispersion is lower than some earlier empirical findings from Serrano (2010)(Ser10), Pakes (1986)(Pak86) or Bessen (2008)(Bes08) that suggest a mean/median ratio between 4 to 11. In these findings, the mean/median ratio is in patents' value. If we enrich the environment and the curvature of

44

profit $\pi'(z)$ is steep, then the dispersion of $v(z)$ will be much larger than the dispersion in z.(See Appendix A.2.12)

2.5.3 Growth decomposition

We define the following growth decomposition by dividing both sides of the law of motion equation (2.5) by z and rearranging terms

$$\lambda \tilde{f}(z) = \alpha \frac{\tilde{F}(z)(1 - \tilde{F}(z))}{z} + \frac{\bar{\mu}(z)}{z} \tilde{f}(z).$$

Integrating over entire z, we have

$$\lambda = \alpha \int \frac{\tilde{F}(z)(1 - \tilde{F}(z))}{z} dz + \int \frac{\bar{\mu}(z)}{z} \tilde{f}(z) dz.$$

If we define $\frac{\bar{\mu}(z)}{z}$ as the innovation rate, then the first term on the right represents the contribution to growth by learning, and the second term is the contribution by innovation. The estimated growth rate λ on the left is 12.86%, and the second term on the right is around 12.75%. This means the direct contribution of innovation to growth is 99%. However, there is an indirect channel that is ignored in this direct calculation: learning opportunities affect the value function through $\alpha \int_z [v(z') - v(z)] f(z') dz$ and provide extra incentives to innovation. The size of this indirect channel depends on the specification of the profit function. In the current setting in which $\bar{\pi}'(z) = z^{\frac{2}{5}}$, this indirect effect is about 0.7% of economic growth.
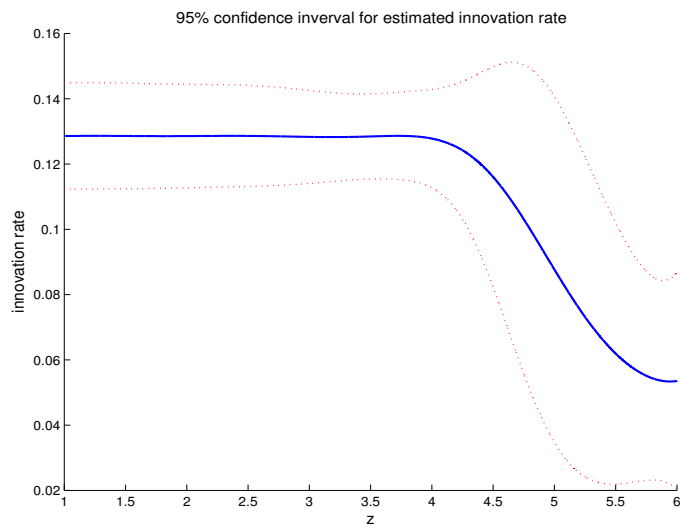
Figure 2.7: Innovation rate.
Estimated innovation rate $\frac{\bar{\mu}(z)}{z}$ is illustrated by the solid blue line and the 95% confidence
interval by the red dotted line.

2.5.4 Inferred equilibrium cost of innovation and value function

We fix $\bar{\pi}'(z) = z^{\frac{2}{5}}$[9] and plug the estimated $\hat{\bar{\mu}}(z)$ to the second step algorithm through
equations (2.6) and (2.7). The second-order derivatives change sign. There is non-concavity
in the value function and non-convexity in the cost function in the agent's level of knowledge.

---

[9]When we adjust $\bar{\pi}'(z)$, the result for other $\bar{\pi}'(z)$ shows revenue function might become flatter, but the
qualitative result holds. One might use other data source to quantify the shape of the $\pi'(z)$
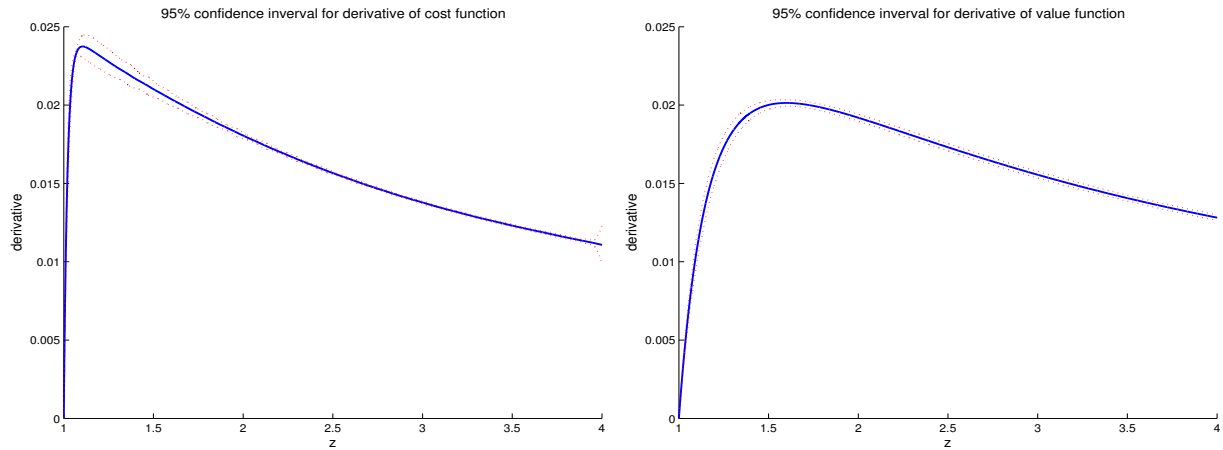
Figure 2.8: First-order derivatives of the cost of innovation and the value function. Estimated derivative of the cost function $\bar{\Psi}(\bar{\mu}(z))$ and value function $v(z)$ is illustrated by the solid blue line and the 95% confidence interval by the red dotted line.

However, this result is sensitive to the initial value of $v'(0)$ and $\bar{\Psi}'(\bar{\mu}(0))$. When $v'(0)$ is larger, we have $v(z)$ is concave in z.

Figure 2.9: First-order derivatives of the cost function and the value function when $v'(0)$ is larger. Estimated derivative of the cost function $\bar{\Psi}(\bar{\mu}(z))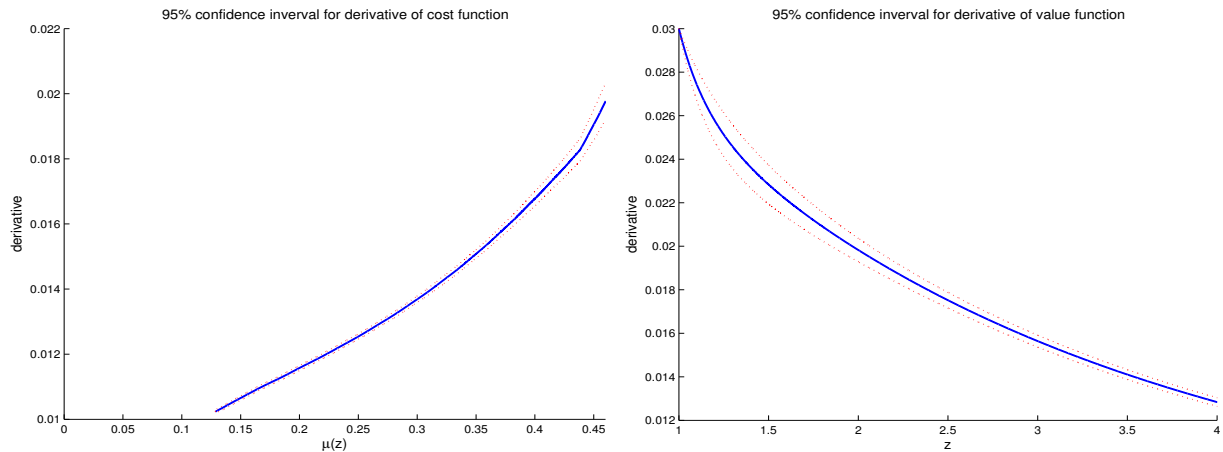$ and value function $v(z)$ is illustrated by the solid blue line and the 95% confidence interval by the red dotted line.

If one needs to identity the entire cost function, one might impose extra exclusion re-

striction and then plug the equilibrium $\hat{\hat{\mu}}(z)$ into the second step of the standard two-step estimator such as B.B.L..

2.5.5   Citations over time

In Abrams et al. (2014)(AAP14), the researchers documented that there is an inverted-U shape relationship between the lifelong patent citations and the patent value. They explained this fact by a non-monotone relationship between the patent value and the knowledge (quality) level in the patent. A higher level patent has more lifelong citations but not necessarily more value. There are two reasons why a patent would have a higher value. First a patent opens a new research area and secondly, a patent significantly increases competitors' entry costs. The first effect receives more citations while the second effect decreases the citations.

We explore another mechanism in which this non-monotone relationship could be explained by a non-monotone relationship between patent quality and patent citations over time. It allows for a monotone relationship between patent quality and value. When a technology is relatively advanced, it can inspire many followers. It is also true that for these followers it is harder to make "real and significant" improvement beyond this technology and create citations. As the economy moves forward, the relative knowledge level in the patent shrinks due to creative destruction. The citation rate m_{ti} will move along the citation rate curve to the left ($m_{ti} = \alpha \tilde{F}(z_i e^{-\lambda t})(1 - \tilde{F}(z_i e^{-\lambda t}))$) over time (in Figure 2.5.5). There are less followers but the probability of making improvement is higher. For patents with higher knowledge levels, this process shows the hump-shaped marginal diffusion rate presumed in the empirical patent literature. In this environment, the cross sectional pattern is the same as the time series pattern. The lifelong citations of a patent also shows a hump-shape in its knowledge level. This holds in our environment especially for a smaller C. For instance, in Figure 2.12, we plotted the cross section citation rate curve in year 2 (thin line), year 5 (dash line) and year 8 (thick line). For patents with level of 3.3, the expected citation rate first increases in year 5 (solid arrow) and decreases in year 8 (dashed arrow). For patents with level of 2.1, the expected citation rate decreases monotonically.
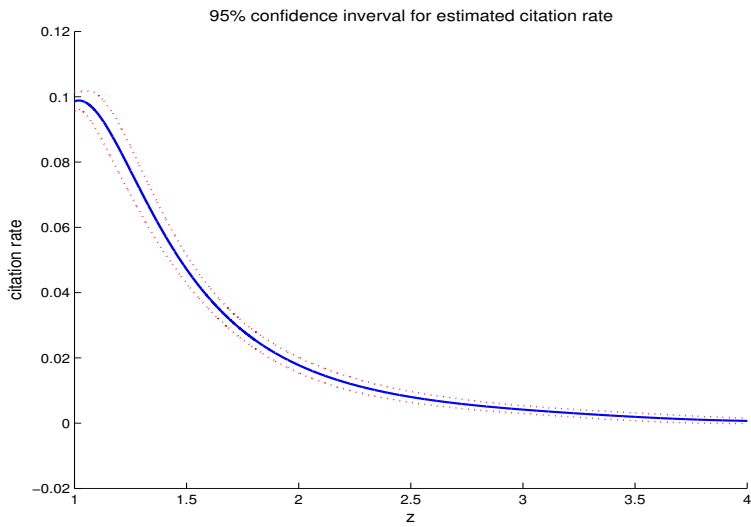
Figure 2.10: Estimated citation rate.

Estimated citation rate in blue solid line, 95% confidence interval in red dash line.
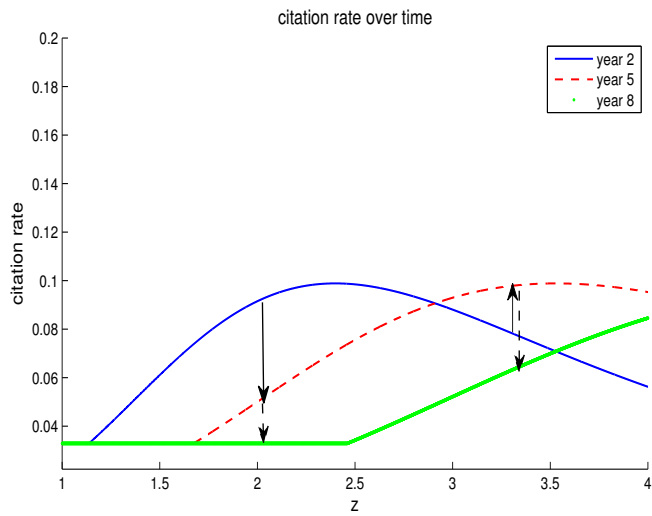
Figure 2.11: Citations received over time.

The citation rate increases first and then decreases for some quality range ($z = 3.3$) over time.

For lower level ($z = 2.1$), the citation rate decreases over time.
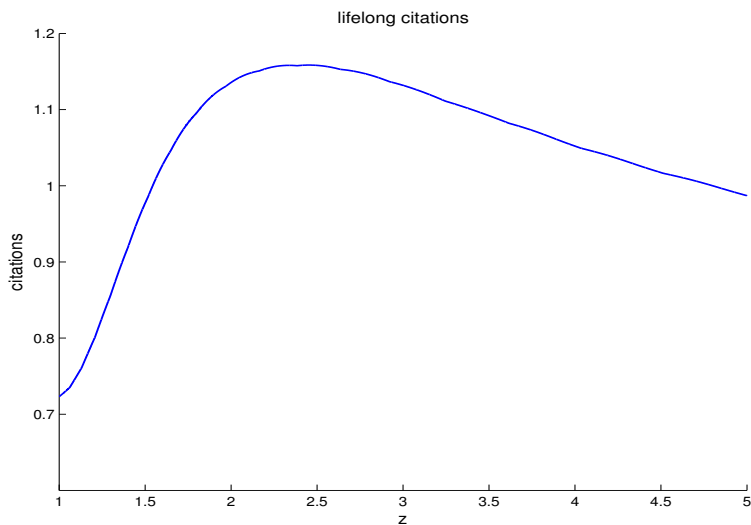
Figure 2.12: Lifelong citations received over time.

The inverted-U shaped lifelong citations for a smaller C.

2.6 Conclusion

Before we drew the conclusion, it worths mentioning the limitation of using patent citation data. There is disincentive for generating patent citations due to financial reasons and non patent referencing such as articles might offer better measurement for learning as argued in Roach and Cohen (2013)(RC13). The role of learning might be underestimated. It would be interesting to compare the results using non patent references in future.

In this paper, we identified and estimated the cost of innovation in an endogenous growth framework with technology diffusion. We also provided suggestive evidence on the relative contributions to growth by learning and innovation.

The two-step algorithm presented in this paper is the first estimator of this class of continuous time macroeconomic models that has recently emerged. This class of models has been used to study the interactions among heterogeneous agents, consumers and firms. They have addressed some important cross-sectional questions regarding wealth distribution across households, capital and employment reallocation across firms and knowledge diffusion

among agents in an economy. Analytical results from these models make comparative statics more tractable.

These models consist of a value function of individual choice (Hamiltonian-Jacobian-Bellman Equation as equation (2.1)) and a law of motion equation (Boltzman, Fokker-planck, Fisher-KPP equation as equation (2.2)), which describes the evolution of the distribution of the state variables. Achdou et al. (2014)(ABL14) has a more detailed review of the standard setups. By studying the inverse problem of each equation, one can identify and estimate deep parameters such as cost function or utility function from the observed macro distribution or reconstruct them non-parametrically. In the appendix, we show similar steps can be applied to the Luttmer (2007)(Lut07) firm dynamic model, the original Lucas and Moll (2014)(LM14) knowledge diffusion model, and an extension of the current model with Brownian motion. Applications to the continuous time finance literature is another area to explore.[10]

The advantage of this approach is the direct identification through reconstruction. It exploits the optimality condition and the relationship between individual choice and the macro distribution. This approach avoids the traditional simulation of the agents to match certain moments in the economy. We show how to reconstruct the objects of interest directly from the data non-parametrically. This allows for a better fitting of the data as well as a flexible structure to accommodate heterogeneity. One of the limitations of this approach is the strong assumption of the large population interaction and the random walk structure. If there is enough micro-level data, direct estimation of the individual policy function can validate the results. A second limitation is the small number of the choice variables that can be reconstructed. With a one dimensional law of motion equation and some further conditions, two elements could be reconstructed in some cases. With a high dimensional law of motion equation, it is relatively easier to reconstruct more elements. A third limitation is that we only consider the distribution and the value function on the balanced growth path. This reduces the partial differential equations to ordinary differential equations. The

---

[10]Entry and exit can be modeled as a separate source term in the law of motion equation of the macro distribution.

uniqueness and existence of the inverse reconstruction is easier to show. The case of the partial differential equation requires more advanced treatments to reconstruct the element. This is an interesting direction for future research.

One important piece of the heterogeneous models in macroeconomics is the dynamic choice problem of the agents where they differ by their states. Those models are often calibrated against macro level data. On the other hand, the identification and estimation of single agent dynamic choice problem have been well studied in empirical IO literature. The difference between the macroeconomic version and the empirical IO version is that in macro models, the distribution of the heterogeneity is an equilibrium object that affects the individual choice, while in empirical IO this information is often discarded by the availability of micro level data. This paper studies the mapping between the individual choice at the micro level and its distribution at the macro level. Under certain conditions this mapping is invertible meaning that one can recover the individual choice from the macro distribution. Then estimation method in empirical IO could be naturally extended to macroeconomic frameworks. This could provide semiparametric or nonparametric identification and inference to heterogeneities in macroeconomic models. Given its similarity to the dynamic discrete choice models, we expect that more insights from empirical IO literature could be utilized in this class of models.

CHAPTER 3

Income and wealth distribution

3.1   Introduction

What drives wealth inequality? Wealth distribution and its dynamics have been a central topic in economics. In this paper, I present a simple framework where the wealth distribution is driven by the precautionary savings of the agents in the incomplete financial market. With some degree of abstraction, this captures the basic features of the consumption saving tradeoff. This is driven by two forces,namely a risky income process and the consumption smoothing motive in the incomplete financial market. Within this framework, I examine which features of the income process and utility function contribute more to the precautionary saving decision and drive the income wealth distribution.

If agents exhibit much heterogeneity in their underlying income processes, the aggregate wealth distribution will be affected. Imagine that the higher income agents have better access to social resources and continue to have a high income profile with relatively higher variances and if the marginal utility of wealth does not fall fast enough, their wealth will continue to grow due to precautionary saving. Thus inequality gap becomes larger. However, if the income process has a strong mean reverting effect and the income uncertainties faced by agents are similar, together with a flat marginal utility function, the wealth inequality will shrink. It is clear from this example that three things matter: the degree of autocorrelation of the income process, the uncertainty associated with the income process, and the relative level and curvature of the utility function. My paper will address these three factors and their empirical importance.

The model will be a continuous time version of Aiyagari (1994)(Aiy94)'s model consisting

of heterogeneous agents with different wealth and income levels. They have access to one risk free asset, namely saving. Saving pays interest at a constant rate. Following Aït-Sahalia (1996)(Ait96)'s approach, the distribution function plus estimation of first moments of the saving and income from the panel data will provide the estimates of implied local variance of the income process. In a second step the utility function could be recovered semi-nonparametrically. The recovered local volatility shows large variations between different wealth and income percentiles.

In the extended model, I also compare it with other formulations of the income process and consider the impact from higher moments. If the income uncertainty is significantly different from the log normal distribution, then incorporating higher moments would be important to better characterize the income process.

The results show that the income process is deviates substantially from the traditional log normal assumption and information up to first three moments of the income process is important in explaining the cross-sectional differences in the saving rate. Risk aversion is not constant and varies across wealth level and consumption.

Previous research has provided some evidence concerning the income process and most recently Guvenen et al (2015)(GKO15) provided a thorough examination of the income process from US social security data and revealed rich heterogeneity of the income process. This is important in understanding the heterogeneity in the saving process. In the paper they examined the life-cycle property and the conditional skewness and kurtosis from the panel structure. They discovered the asymmetric shift of the skewness of the labor income in the business cycles. My paper focuses on how these moments affect the joint wealth income distribution. Using similar data, Schimdt (2014)(Sch15) proposed an affine model in relating the income risk and some other indexes. My work is closely related in the sense that I will examine the income process and their moments implications for the precautionary saving.

In the incomplete financial market, saving is a primary tool to hedge against income and other sources of shocks. Wealth consequently correlates closely with the income process. Bewley (1977)(Bew77), Hugget (1993)(Hug93) and Aiyagari (1994)(Aiy94) models started

this line of research. Castañeda, Díaz-Giménez and Ríos-Rull (2003)(CDR03) calibrate a detailed model using a similar dataset and target the Lorenz curve of income and wealth distribution. Rather than targeting the percentiles, I will show how the income process might contribute to the cross sectional wealth difference through savings using the income wealth distribution as input. The continuous time modeling was adopted from the continuous time macroeconomics frameworks initiated by Lucas and Moll (2014)(LM14) and Lions and Larsy (2006)(GLL). [1] It allows sharper characterization of the transitional process.

It is worth mentioning that several other forces have been proposed to explain both the inequality and its trend. Among them, entrepreneurship is influential. Business owners have different incomes than employees. They can utilize the ownership as a saving tool. Cagetti and De Nardi (2005)(CD05) studied the career choice and income distribution of workers and firm owners. Quadrini (2000)(Qua00)'s survey paper and recently Buera's (2009)(Bue09) survey paper provided summaries along this direction. These authors also suggest that high saving is associated with entrepreneurship behavior, either by initial investment or as collateral due to borrowing constraints. Meanwhile, some of the risks might be shared by the employees as well. (Zhang's (2014)(Zha14) provides an illustration).

Different levels of wealth may also mean different capability in acquiring financial assets for hedging or investment purposes. Wealthier families might benefit from more complicated asset portfolios and have much lower total risks. My paper does not address career choice or access to financial market directly, but the the process and distribution will be dependent on the wealth level. Finally, bequest motive is also beyond the discussion of this paper.

The paper is organized as follows: In section 3.2, I will present the simple baseline model and the estimation procedure. Section 3.3 provides a summary of the data and estimates the baseline model, while section 3.4 describes some extensions. Section 3.5 estimates the utility. Section 3.6 calculates the risk premium. Section 3.7 concludes.

---

[1] see Achdou et al. (2014)(ABL14) for a detailed description.

3.2 A saving problem

3.2.1 Model

I adopt the continuous time framework for its convenience in exhibition. I start with the baseline model where income is log normal. In continuous time, this means that the income follows geometric Brownian motion. This log normality assumption has been adopted in many previous analysis for its tractability.

Agents receive salaries at rate wz_t where w is the economy-wide wage rate and z_t is the productivity (efficient labor unit) of the agent in the period. Since w would be normalize to 1, z also represents the labor income. The income evolves according to:

$$dz_t = \mu(a_t, z_t)z_t dt + \sigma(a_t, z_t)z_t dW_t, \tag{3.1}$$

where $\mu(a_t, z_t)z_t$ is the drift of the productivity changes and $\sigma(a_t, z_t)z_t$ is the standard error of the volatility experienced by the agent with wealth level a_t and income level z_t. This models that each agent will experience some shocks to productivity and resulting changes to income flow. It is important here that the drift term and the volatility term both depend on the net wealth level. In Buera (2009)(Bue09) and Cagetti and De Nardi (2005)(CD05), they examined the channel where wealth affects the entrepreneurship choice through the borrowing constraint. Then different career implies different income level. Although I did not take the career choice into direct consideration, this formulation potentially captures how different wealth levels might affect income opportunities.

Agents can only save their income at rate r given by the market. The saving/net wealth a_t thus follows:

$$da_t = (w_t z_t + ra_t - c_t)dt, \tag{3.2}$$

where $z_t + ra_t - c_t$ represents the labor income and interest earned minus consumption. We define $s(a_t, z_t) \equiv (z_t + ra_t - c_t)$.

If we denote the ^ variables as the log transform of the original variables, then, for example

$\hat{z} \equiv log(z)$ and $\hat{a} \equiv log(a)$.[2] With a bit abuse of the notation, $\mu(.,.)$ and $\sigma(.,.)$ will represent the log case as the following :

$$dz\hat{} \quad = [\mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma^2(\hat{a}, \hat{z})]dt + \sigma(\hat{a}, \hat{z})dW \tag{3.3}$$

$$d\hat{a} \quad = \frac{s(\hat{a}, \hat{z})}{exp(\hat{a})}dt \tag{3.4}$$

Agents are ex ante identical except for their wealth and income levels. They maximize their expected utility function $\mathbf{E} \int_0^\infty e^{-\rho s}u(c_s)ds$ by choosing the consumption and saving given the income and net wealth, where u is the flow utility and ρ is the discount factor. Using Ito's lemma, the Hamiltonian-Jacobian-Bellman (H.J.B.) equation equation of the value function becomes

$$\rho v(\hat{a}, \hat{z}) = \max_{c_t} \quad u(c_t) + \partial_a v(\hat{a}, \hat{z})s(\hat{a}, \hat{z}) + \partial_z v(\hat{a}, \hat{z})[\mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma(\hat{a}, \hat{z})^2] \tag{3.5}$$

$$+ \frac{1}{2}\partial_{zz}v(\hat{a}, \hat{z})\sigma(\hat{a}, \hat{z})^2.$$

Using the framework of Meanfield games by Lasry and Lions (2006)(GLL) and Lucas and Moll (2014)(LM14) the law of motion equation could be summarized by the Kolmogorov forward equations as

$$0 = -\frac{\partial}{\partial \hat{a}}[g(\hat{a}, \hat{z})\frac{s(\hat{a}, \hat{z})}{exp(\hat{a})}] - \frac{\partial}{\partial \hat{z}}[g(\hat{a}, \hat{z})(\mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma^2(\hat{a}.\hat{z}))] + \frac{1}{2}\frac{\partial}{\partial \hat{z}^2}[g(\hat{a}, \hat{z})\sigma^2(\hat{z})], \tag{3.6}$$

where $g(.,.)$ is the joint distribution. In the stationary case, the left hand side is the derivative to the time change and would be 0.

Before introducing the equilibrium, I will explain the meaning of this law of motion equation. The change in the density consists of two parts, as follows: the change in wealth through saving (the first term) and the stochastic change in the income (the last two terms). These should cancel out in the stationary case. In the discrete case, one encounter the law

---

[2]There were not many agents with negative netwealth in the data and thus they were dropped; we can set the borrowing constraint at a positive number very close to 0.

of motion equation in the form $g_{t+1} = Tg_t$, where subscript t denotes the time period. T is the transitional matrix. Intuitively, one can expand g_t to the second order using g_{t+1} as the base. This results in the law of motion equation (3.6).[3].

The first order condition with c becomes:

$$u'(c_t) = -\frac{1}{exp(\hat{a})}\partial_a v(\hat{a}_t, \hat{z}_t)\partial_c s(\hat{a}_t, \hat{z}_t) \tag{3.7}$$

In discrete case the F.O.C. would not bind for some low net wealth ($\underline{a}$) agents, since they become borrowing constrained. In the continuous type case, the F.O.C. always holds.

3.2.2 Equilibrium

Equilibrium definition: A stationary equilibrium is defined by $r, \hat{a}$, and the distribution $g^*(\hat{a}, \hat{z})$, and optimal saving rule $s^*(\hat{a}, \hat{z})$.

Given r, a, and z, $s^*(\hat{a}, \hat{z})$ maximizes the utility function, and given the optimal $s^*(\hat{a}, \hat{z})$:

$$r = \partial_K F(K, 1), \tag{3.8}$$

where

$$K = \int_{\underline{z}}^{\bar{z}} \int_{\underline{a}}^{\bar{a}} exp(\hat{a})g^*(\hat{a}_t, \hat{z}_t), \tag{3.9}$$

and the joint density $g^*(.,.)$, satisfies the Kolmogorov forward equation (3.6).

3.2.3 Estimation procedure for the baseline model

The estimation goal is to assess the features of the income process that are important in explaining the joint income wealth distribution. In the baseline model, the key features of the income process are the conditional mean and conditional volatility. I will recover the volatility suggested by (3.6) using the empirical conditional mean and empirical joint income wealth distribution. Then, this recovered volatility will be compared with the empirical

---

[3] One might intend to expand g_{t+1} to the second order using g_t as base. However, if time runs forward, future states are unknown and today's states are known. Thus, one would expand today's distribution rather than tomorrow's distribution. Details can be found in section 3.4

conditional volatility in the data. If the recovered volatility is in line with the data, then the income wealth distribution can be explained by heterogeneity in the income process in this simple environment.

This method was proposed by Aït-Sahalia (1996)(Ait96) to estimate the option volatility and option prices. In his paper, he used the panel data to estimate the drift. Estimated drift and marginal distribution were used in the law of motion equation to recover the implied volatility. In the current macroeconomics setting, this approach can single out the variance term directly without fully solving the model. This method also uses the entire distribution from the data rather than first or second moment only. In addition this inverting process will reveal dependence structure between the income process and wealth.

Suppose we have $s(\hat{a}, \hat{z}), g(\hat{a}, \hat{z})$, and $\mu(\hat{a}, \hat{z})$; then we can solve for $\sigma(\hat{a}, \hat{z})$ by integrating both sides of (3.6) once with $\hat{z}$. It then becomes a first order differential equation of $g(\hat{a}, \hat{z})\sigma^2(\hat{a}, \hat{z})$ for fixed $\hat{a}$. Then, dividing the result by $g(\hat{a}, \hat{z})$ will give us $\sigma^2(\hat{a}, \hat{z})$.

In Aït-Sahalia's original approach, he fits the parametric assumption on the drift $\mu(\hat{a}, \hat{z})$ and then plug the estimates into (3.6). Instead, I will estimate those elements nonparametrically to capture richer dependence structures in the data.

The idea behind this different approach was proposed in the seminal paper of Johannes (2003)(Joh04) (or see Gihkman and Skorohod (1972)(GS72)) in estimating the interest rate.
4

$$\frac{1}{\Delta}E(d\hat{z}|\hat{a}, \hat{z}) \quad = \mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma^2(\hat{a}, \hat{z}), \tag{3.10}$$

$$\frac{1}{\Delta}E(d\hat{a}|\hat{a}, \hat{z}) \quad = \frac{s(a, z)}{exp(\hat{a})}, \tag{3.11}$$

$$\frac{1}{\Delta}E((d\hat{z})^2|\hat{a}, \hat{z}) \quad = \frac{2}{\Delta}E(d\hat{z}|\hat{a}, \hat{z})^2 + \sigma^2(\hat{a}, \hat{z}). \tag{3.12}$$

The left handside expectations could be replaced by their empirical counterparts with

---

[4]The first term on the right hand side of the last equation is different from that in the original paper because the data have a longer horizon; thus $d\hat{z}$ could not be ignored. This could be derived through a Taylor expansion or using Dynkin's formula.

local nonparametric regression equations:[5] I will plug the nonparametric estimates (3.10) and (3.11) back into (3.6) and calculate the recovered local volatility $\sigma_{im}^2(\hat{a}, \hat{z})$ and then compare it with the nonparametric estimates in (3.12). If the two correlate well, then the model provides a reasonable framework in understanding in the income process and income wealth joint dynamics.

3.3 Data and estimation

3.3.1 Data description

The data comes from the Survey of Consumer Finance (SCF). One of the recent survey, includes a panel data structure from 2007 to 2009. I will extract the income (z), saving/ net wealth (a) then form the joint density $g(\hat{a}, \hat{z})$ of the net wealth and income. The total income is defined as the sum of primary earnings plus secondary earnings. The difference between total assets and debts is defined as the net wealth, where total assets includes liquid assets, home assets, business assets, lending, vehicles, IRA , life insurance, misc assets and Pension1. Total debts includes credits, loans and mortgage.

There are around 12,000 observations with sampling weights after imputation. I use the nonparametric method to estimate the densities. The nonparametric kernel density estimator of the marginal density of m for $m = \hat{a}, \hat{z}$ is defined as $g(m) = \sum \frac{w_i}{sum(w_i)} K\left(\frac{m-x_i}{h}\right)$ where K is the gaussian kernel, h is the bandwith, and x_i is the data. The bandwith is chosen by the rule of thumb $\frac{4}{3}^{\frac{1}{5}} \sigma n^{-\frac{1}{2r+1}}$ and $r = 3$ prepared for later calculations. The joint density is similarly defined using the product kernel.

The estimated results are as shown in Figure 3.1 and Figure 3.2.

---

[5]Bandi and Phillips (2003)(BP03) and Bandi and Nguyen (2003)(BN03) provides more detailed treatments using local time approach which provides similar estimation methods based on ergodicity.

3.3.2 Labor income
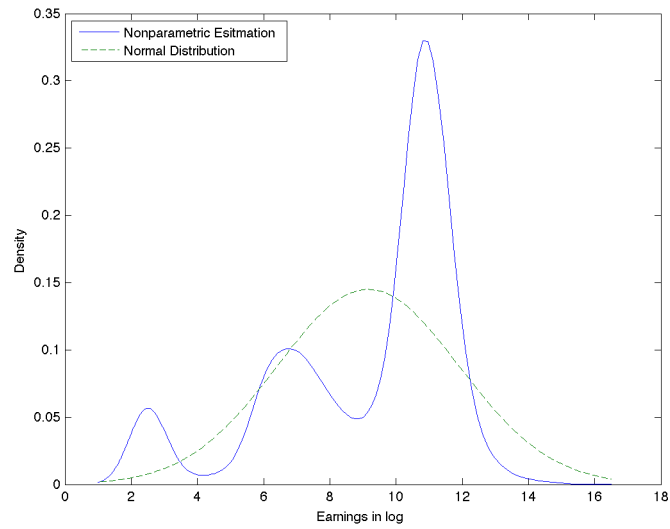
Figure 3.1: Income distribution.

Name	log earnings
Mean	9.158
Standard deviation	2.752
Skewness	-1.068
Kurtosis	3.276

Table 3.1: Statistics of the log total income.

percentile	Log income	Income
10	5.64	280
25	7.244	1,400
50	10.404	32,991
75	11.082	64,991
90	11.6	10,910
95	11.931	15,190
99	12.96	425,070

Table 3.2: Percentiles of the log total income, income level in parentheses.

3.3.3 Net wealth
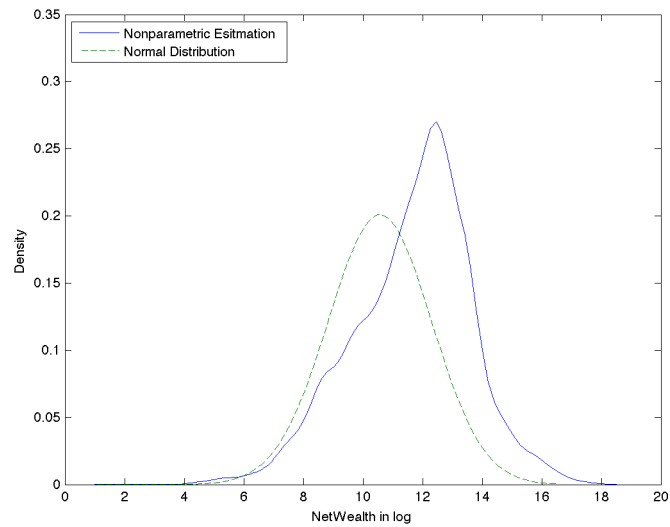
Figure 3.2: Net wealth distribution.

Name	Log net wealth
Mean	11.644
Standard deviation	1.985
Skewness	-0.3940
Kurtosis	3.345

Table 3.3: Statistics of the log net wealth.

Percentile	Log net wealth	Net wealth
10	8.853	7,000
25	10.404	33,000
50	11.926	151,100
75	12.938	415,800
90	13.825	1,009,500
95	14.602	2,195,700
99	16.032	9,175,100

Table 3.4: Percentiles of the log net wealth, income level in parentheses.

In both graphs, data are in the blue line. The green line represents the normal distribution with the same mean and variance. Both log income and log net wealth skewed to the right. The log net wealth shows a fat tail. Both log earnings and log net wealth shows a negative skewness and a kurtosis greater than 3. This negative skewness and excess kurtosis have been documented in the Guvenen et al. (2015)(GKO15). Castañeda, Díaz-Giménez and Ríos-Rull (2003)(CDR03) also used a stochastic efficiency labor framework to capture income volatility. In their calibration, there were four levels of efficiency namely 1, 3.15, 9.78, 1,061, and the distribution decreased monotonically. In their calibration, the skewness was 1.0536 and kurtosis was 3.312 [6]. The skewness in their finding was positive, which contrasted with recent findings. This might have occurred their calibration target is the inequality Gini coefficient and four quartiles rather than the transitional probability.

---

[6]See Table 4 and Table 5 of their paper and transform to log efficiency labor unit.

3.3.4 Estimation of the income process and saving rate

We start from the baseline model where income z follows a Geometric Brownian motion. Then, we will enrich the process and show how this translates into saving behavior.

As mentioned previously, I will estimate the conditional first and second moments of $d\hat{z}$, and the conditional first moment of $d\hat{a}$, and implied variance $\sigma^2_{im}(\hat{a}, \hat{z})$.

We define the saving rate by savings denominated by the net wealth (e.g., $\frac{s(\hat{a},\hat{z})}{exp(\hat{a})}$), which is the first moment of the change in log net wealth estimated in (3.11). The results are listed in Figure 3.3. Overall, the saving rate is very low for most of the agents. It decreases with the wealth level. However, during 2007-2009, most of the agents experienced the financial crisis, lost assets and reduced their savings. Within the wealth percentiles, the higher-income agents tend to save more.
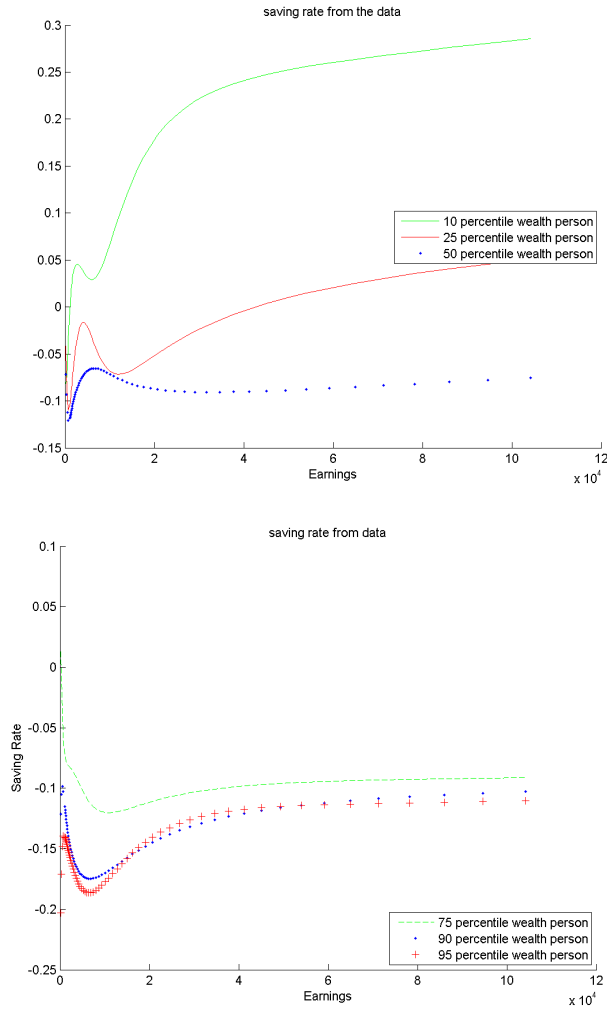
Figure 3.3: Saving rates in the data.

In Figure 3.4, I plot the estimated first moment of $d\hat{z}$, and the change in the log income. Comparing to the calibration results in Castañeda, Díaz-Giménez and Ríos-Rull (2003)(CDR03)(in the right panel), where their income process only conditions on past income level, my estimates show a similar pattern up to around 50th percentile income level. Their calibrates show a different pattern for the high income agents.
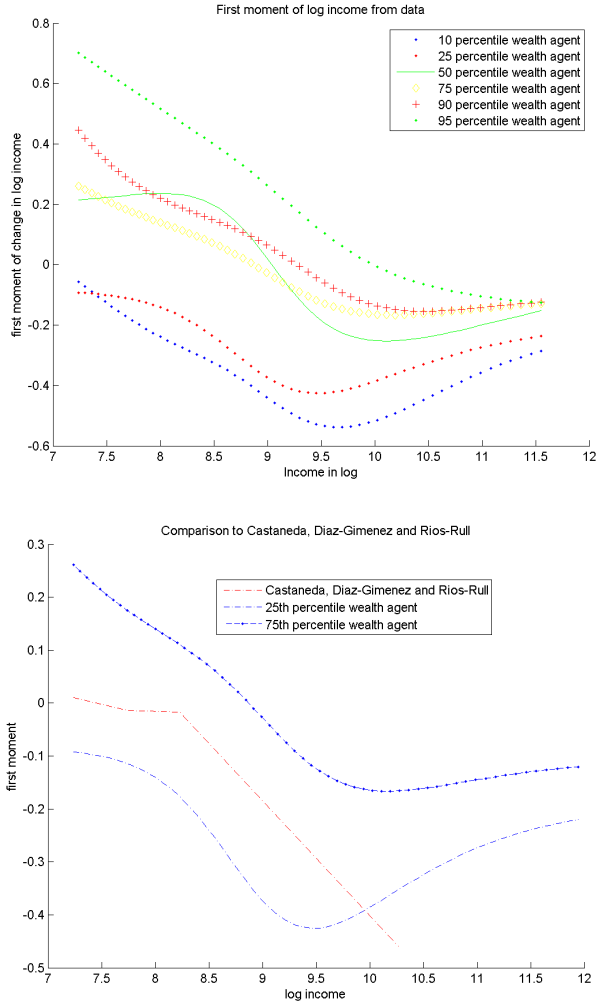
Figure 3.4: First moment of $d\hat{z}$.

Figure 3.5 depicts the drift term $\mu(\hat{a}, \hat{z})$. For income levels below 13,360 there is very high heterogeneity. The 95th percentile wealth agents have a higher income growth. For higher income agents the mean reverting force is weaker and there is diverging income inequality.
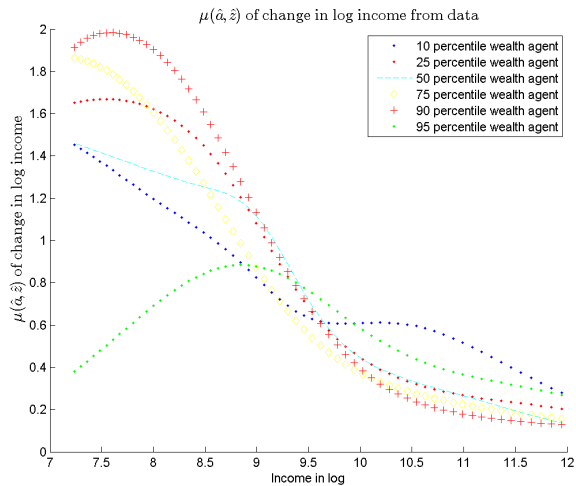
Figure 3.5: $\mu(\hat{a}, \hat{z})$.

In the following panels of Figures 3.6-3.7, I plot the implied volatility $\sigma_{im}(\hat{a}, \hat{z})$ (green dot) derived from the Kolmogorov equation, data (red line), and constant parametric estimation (blue dot).

Before showing these figures, what should be expected? If income process and its volatility could be fully captured by the first and second moments, then the recovered volatility calculated by (3.6) should be very close to the second moments of $d\hat{z}$: $\sigma^2(\hat{a}, \hat{z})$.

67
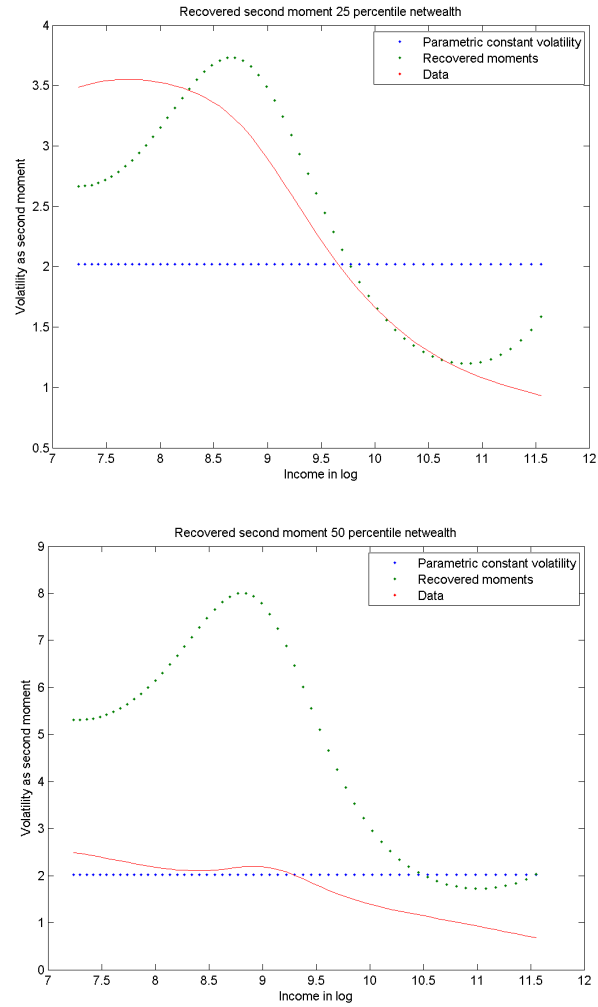
Figure 3.6: Second moment of $d\hat{z}$ and implied volatility part 1.
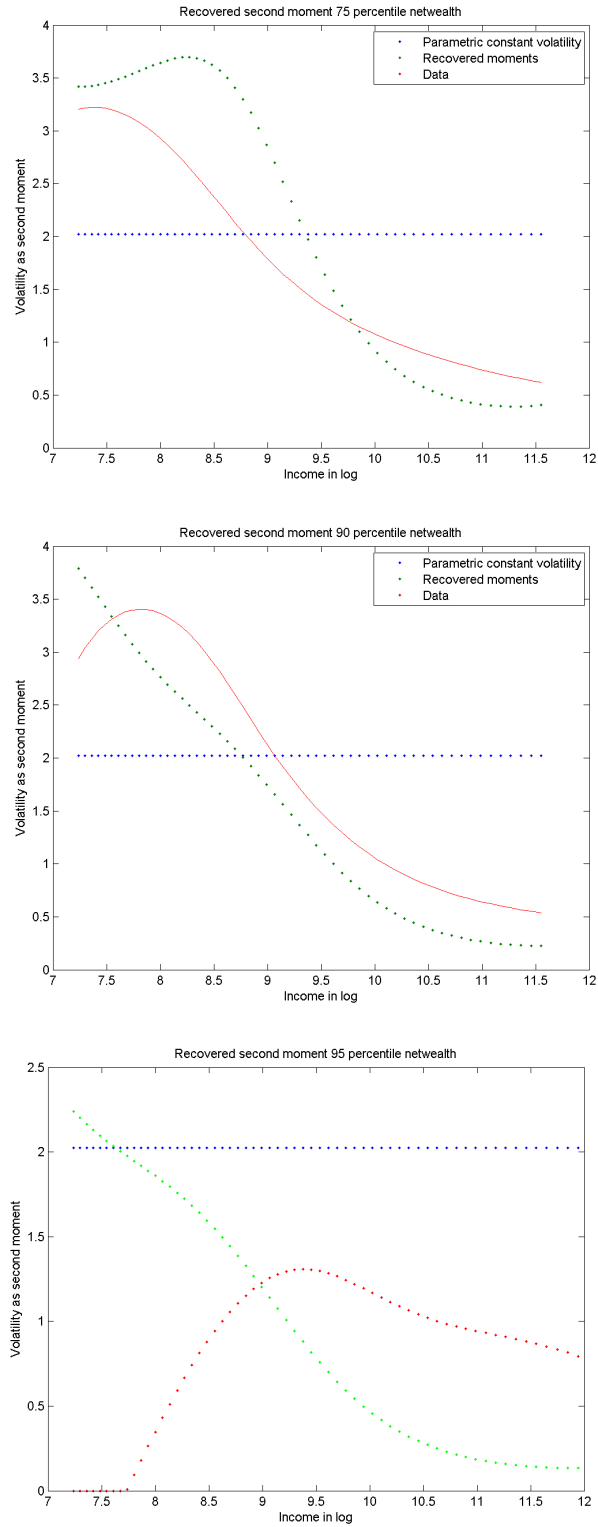
Figure 3.7: Second moment of $d\hat{z}$ and implied volatility part 2.

As shown in the graphs, the constant volatility estimation tends to capture the case for mid income agents. It overestimates the income volatility of high income agents and underestimate that of lower income agents.

The baseline model is able to reproduce the variance of the log income for relatively lower wealth and higher wealth agents. It follows the data closely for the 25th and 90th percentile wealth agents and is reasonably close for 75th percentile wealth agents. For the super wealthy agents, the baseline model does not produce enough volatility compared to the data. Since the normal distribution does not have strong fat tail, it might have poor prediction power to the rich high income agents.

Castañeda, Díaz-Giménez and Ríos-Rull (2003)(CDR03) obtained an increasing conditional standard deviation, which supports the reasoning that high income agents have more volatility and save more. The nonparametric estimates, however, show a decreasing volatility. Similarly, if we examine the conditional skewness of the income, the level is roughly consistent for the highest efficiency level agent, though the cross sectional trend is different from the data. The Guvenen et al. (2015)(GKO15) estimates also show a decreasing trend for the standard deviation.
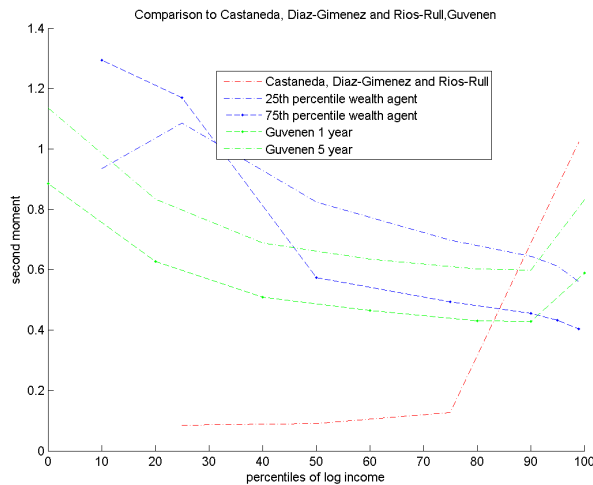
Figure 3.8: Comparison with previous studies.

3.4 Implication of higher moments

As shown in the previous section, the local log normal distribution neglected to capture some features of volatility in the income process. In particular, it did not generate enough volatility for high income rich agents and created too much volatility for the mid wealth agents. This led to the consideration of relaxing the log normal structure and incorporating higher moments. If large income shocks occur with a greater probability, this will be captured by the higher moments.

Recently, many authors in the literature have suggested that the distribution of log income changes has nonnormality feature that affects the skewness and kurtosis of the income process significantly. Guvenen et al. (2015)(GKO15) and Schimdt (2014)(Sch15) both studied this effect using the Social security data. Although the SCF data has a shorter panel with a lower frequency, it still generates some of these features.
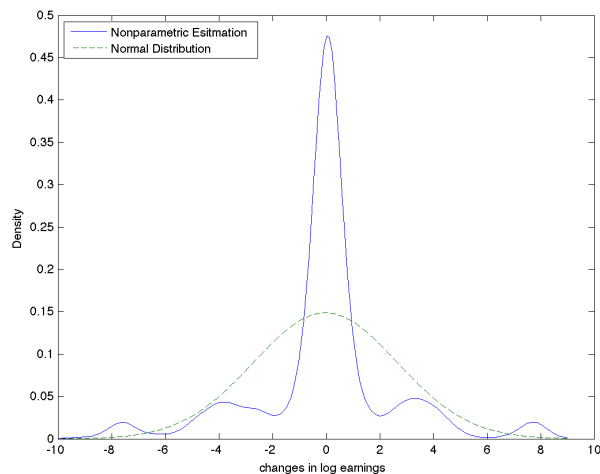
Figure 3.9: Transitional density of $d\hat{z}$

If we examine the unconditional transitional density of the change in log income (shown in Figure 3.9) and compare it to a normal distribution with the same mean and variance, we find sharp differences between the two. The change in log income shows a much fatter tail in both tails. It also generates higher kurtosis than a normal distribution. In Appendix A.3.1, I present the comparison with estimates from Guvenen et al. (2015)(GKO15) on skewness

and kurtosis.

I introduce two formulations to incorporate the effects of higher moments. The first adds more moments into the transitional process. In the Brownian motion setup, the Kolmogorov forward equation only contains the first and second moment of the income process. In the non-normal case more moments could be incorporated into the Kolmogorov forward equation and examined for their effects on the saving. Details are discussed in section 3.4.1.

The second alternative is the jump process. Income sometimes experiences larger deviations at a much lower frequency. This increases moment values at the tail and generates a greater impact on the saving rate because the change in marginal utility is bigger.

3.4.1   Higher-order effects

In this section, I provide an alternative framework which utilizes the moment conditions $E((d\hat{z})^n|\hat{a}, \hat{z})$. The probability density $g(x)$ follows:

$$g(x, t + \Delta t) = \int T(x, t + \Delta t|x', t)g(x', t), \tag{3.13}$$

T is a transitional density from x'[7] at t to x at $t + \Delta t$. If we have all the moment conditions of the transitional density, then we can use inverse Laplace (Fourier) transform to derive the transitional density. With some manipulation,[8] we have:

$$\frac{\partial g(x)}{\partial t} = \sum_{n=1} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial x^n}[(M_n(x, t)g(x)], \tag{3.14}$$

where $M_n(x, t)$ stands for the conditional moments for the transitional densities and is defined by $M_n(x, t) = \int (x - x')^n T(x, t + \tau|x', t)dx'$. The law of motion equation used in section (3.2) is a special case of this expansion only up to two moments.

Setting the stationary condition $\frac{\partial g(x)}{\partial t} = 0$ and substitute in the moment conditions for

---

[7]Here x is a vector containing log net wealth and log income.
[8]see Risken (1978)(Ris78) for details

changes in log income, we have:

$$\frac{\partial}{\partial \hat{a}}[g(\hat{a}, \hat{z})\frac{s(\hat{a}, \hat{z})}{exp(\hat{a})}] = \sum_{n=1} \frac{(-1)^n}{n!} \frac{\partial^n}{\partial \hat{z}^n}[E((d\hat{z})^n|\hat{a}, \hat{z})g(\hat{a}, \hat{z})]. \qquad (3.15)$$

To illustrate this effect, we can integrate both sides with respect to $\hat{a}$ and calculate the contribution of each term:

$$\frac{s(\hat{a}, \hat{z})}{exp(\hat{a})} = \sum_{n=1} \frac{(-1)^n}{n!} \int^{\hat{a}} \frac{\partial^n}{\partial \hat{z}^n}[(E(d\hat{z})^n|\hat{\tau}, \hat{z})g(\hat{\tau}, \hat{z})]d\hat{\tau}/g(\hat{a}, \hat{z}). \qquad (3.16)$$

This shows how higher-order moments affect the saving rate. Since we truncate the infinite expansion using finite terms, we introduce the fitting coefficients $\beta_n(\hat{a})$. $\underline{a}$ is fixed at 25th percentile:

$$\frac{s(\hat{a}, \hat{z})}{exp(\hat{a})} - \frac{s(\underline{\hat{a}}, \hat{z})}{exp(\underline{\hat{a}})} = \beta_0 + \sum_{n=1} \beta_n(\hat{a})\frac{(-1)^n}{n!} \int_{\underline{a}}^{\hat{a}} \frac{\partial^n}{\partial \hat{z}^n}[(E(d\hat{z})^n|\hat{\tau}, \hat{z})g(\hat{\tau}, \hat{z})]d\hat{\tau}/g(\hat{a}, \hat{z}) \qquad (3.17)$$

First, I will present the results using only the first two moments without adjusting the coefficients.
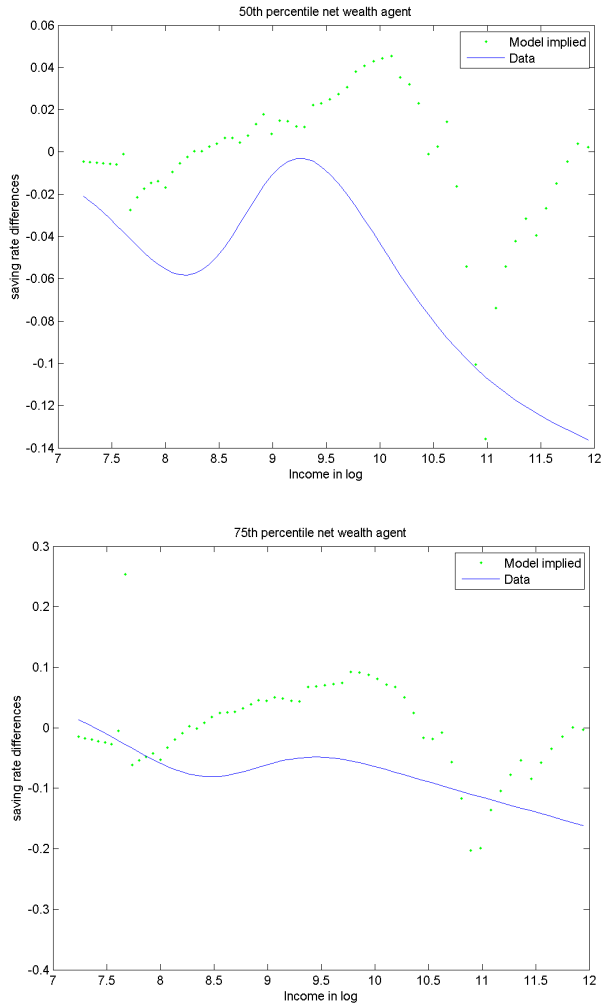
Figure 3.10: Implied saving rate using the first two moments without fitting part 1.
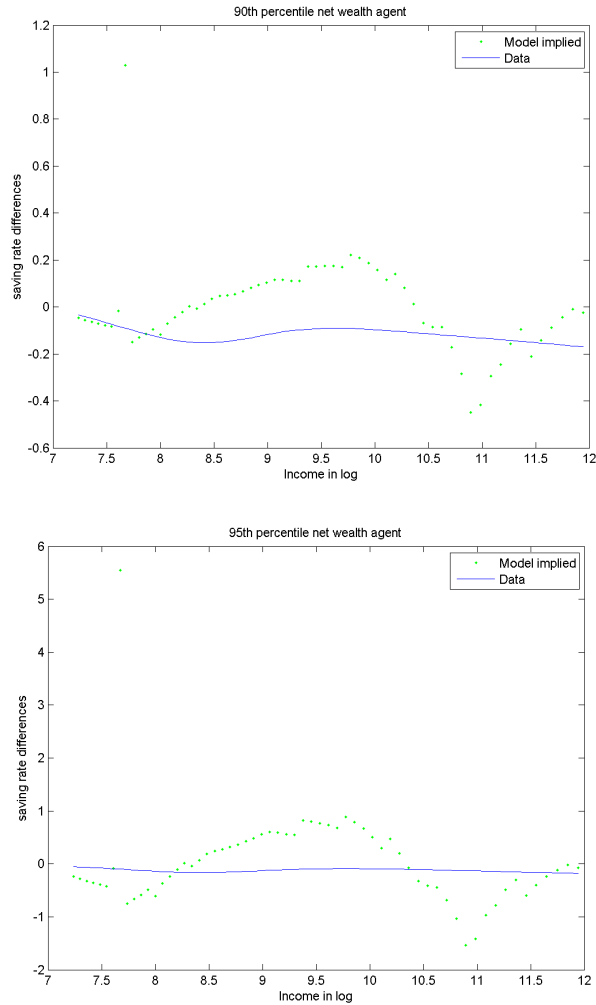
Figure 3.11: Implied saving rate using the first two moments without fitting part 2.

Shown in the above two panels, the blue line is the difference in the saving rate between the target percentile net wealth agent and saving rate of the 25th percentile agents. The green dotted line is that suggested by the model using only two moments, without the fitting coefficients. The figures clearly shows that only two moments [9] are not sufficient to explain the cross-sectional difference in the saving rate.

In Figure 3.12-3.13, I present the results of fitting in higher moments to predict the difference in the saving rate. The blue line is the saving rate data, and the light blue dotted

---

[9]if the model contains only two moments like the log normal baseline, then there should be no fitting coefficients.

line represents using only first moments. The green dotted line represents using first two moments. The red dotted line represents using first three moments, and yellow dotted line represents using the first four moments.
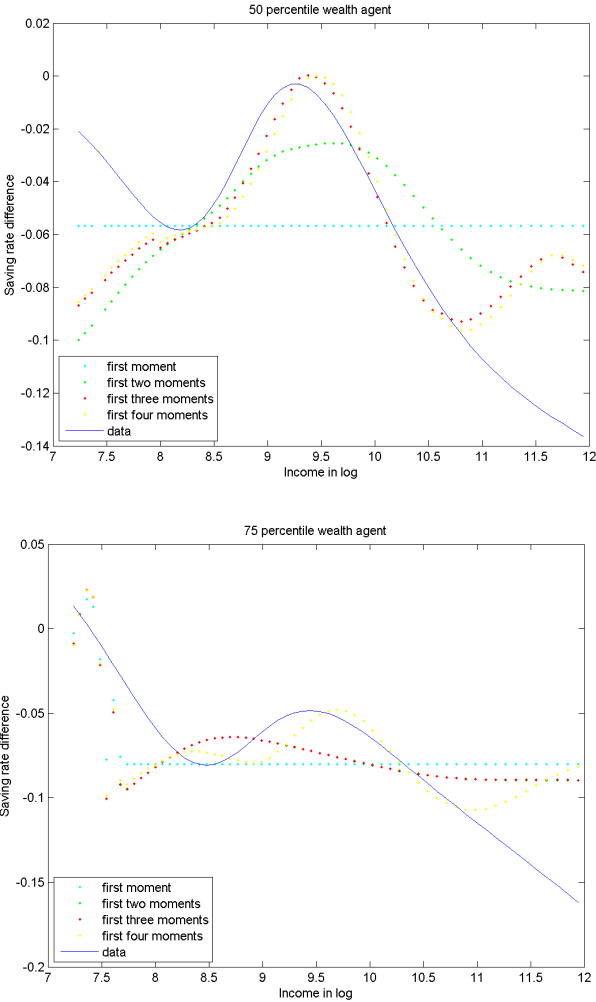
Figure 3.12: Comparison different percentile net wealth agent part 1.
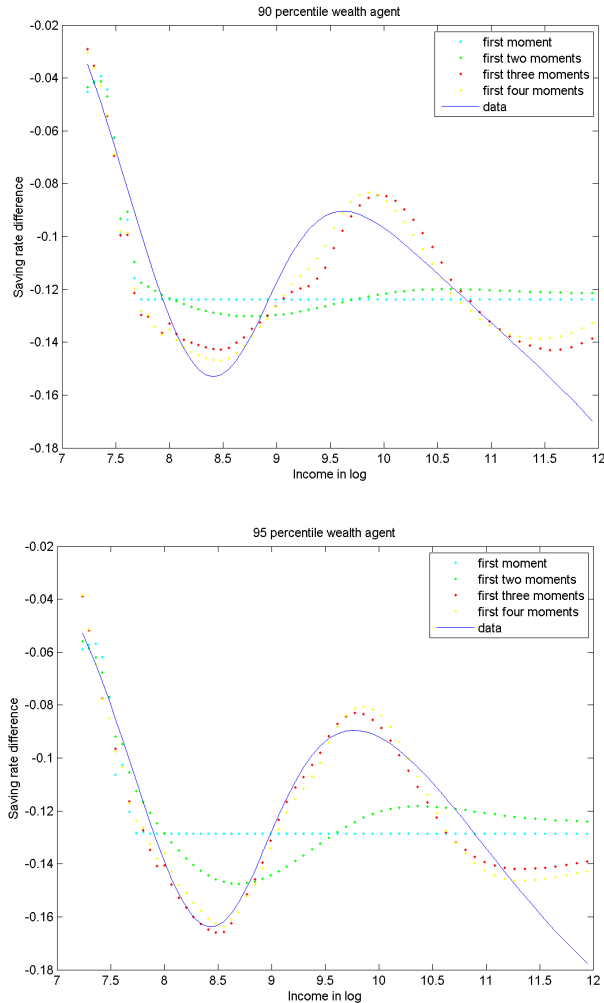
Figure 3.13: Comparison different percentile net wealth agent part 2.

In Figure 3.6-3.7, we found that implied variance was far off for the 50th and 75th percentile net wealth agents. It is clear that the first two moments did not fully capture the income process for these agents. Adding in higher moments, especially the third moment, significantly improve the fitting. Overall, the first three moments captured important features of the income process to reproduce the saving rate, except for 75th percentile net wealth agents, where the fourth moment became important.

Indirect evidence could be obtained by regressing the saving rate on the empirical moments directly. In Table 3.5, I provide the regression results on the moments. This result is based on the pooled five imputes. Although significant in the pooled data, the second

moment becomes insignificant in some replicas.

Variable	Coefficient
	(t stat in parenthesis)
d	0.142
	(8.87)
$[d\hat{z}]^2$	0.026
	(3)
$[d\hat{z}]^3$	-0.016
	(-5.83)
$[d\hat{z}]^4$	-0.005
	(-4.51)
$[d\hat{z}]^5$	0.001
	(5.53)
$[d\hat{z}]^6$	0.0002
	(5.07)
$[d\hat{z}]^7$	-0.000015
	(-5.67)
$[d\hat{z}]^8$	$-3.07 \cdot 10^{-6}$
	(-5.25)
$[d\hat{z}]^9$	$9.46 \cdot 10^{-8}$
	(6.06)
$[d\hat{z}]^{10}$	$1.64 \cdot 10^{-8}$
	(5.37)
Intercept	-0.198
	(-18.77)

Table 3.5: Dependent variable: saving rate $\frac{s(\hat{a},\hat{z})}{exp(\hat{a})}$

3.4.2 A jump diffusion formulation

In this section, I will estimate the income process using the jump diffusion structure. The jump follows a poisson process at rate $\lambda(\hat{a}, \hat{z})$, where jump sizes ξ satisfies the double

exponential distribution :

$$p_1\eta_1 e^{\eta_1 y}\mathbf{1}_{y<0} + (1-p_1)\eta_2 e^{-\eta_2 y}\mathbf{1}_{y>0}.$$

This contains two asymmetric parts. If jumps upward, it follows exponential distribution with intensity parameter (η_2). $1 - P_1$ is the probability of jumping upwards. Similarly η_1 and p_1 is the parameters for jump down. Here, those parameters depend on $\hat{a}, \hat{z}$.

The identification method is the same as the previous section where local conditional moments were used:

$$\frac{1}{\Delta}E(d\hat{z}|\hat{a},\hat{z}) = \mu(\hat{a},\hat{z}) - \frac{1}{2}\sigma^2(\hat{a},\hat{z}) + \lambda(\hat{a},\hat{z})E(\xi)$$

$$\begin{aligned}
\frac{1}{\Delta}E((d\hat{z})^n|\hat{a},\hat{z}) &= nE((d\hat{z})^{n-1}|\hat{a},\hat{z})(\mu(\hat{a},\hat{z}) + \frac{1}{2}\sigma^2(\hat{a},\hat{z})) \\
&\quad + \frac{n(n-1)}{2}E((d\hat{z})^{n-2}|\hat{a},\hat{z})\sigma^2(\hat{a},\hat{z}) \\
&\quad + \lambda(\hat{a},\hat{z})(E(dz+\xi)^n - E(\xi)^n) \\
E(\xi^n) &= p_1 k!(-1)^k\eta_1^{-k} + (1-p_1)k!\eta_2^{-k}
\end{aligned}$$

Jump intensity and variance of jump sizes could be recovered using these relationships.

This distribution has been used previously by Kou (2002)(Kou02) and a similar version with the double Pareto distribution was proposed in Moll, Kaplan and Violante (2015) (MKV15). The double exponential distribution will generate a sharp shape in the middle part of transitional density. The asymmetric setup allows for different intensities for different jump directions.

Figure 3.14-3.16 illustrates the decomposition of the estimated second moment of the income into diffusion risk and jump risk. The red dotted line represents the jump risk $\lambda(\hat{a},\hat{z})Var(\xi)$. The blue line is the second moment which is the sum $\sigma^2(\hat{a},\hat{z})+\lambda(\hat{a},\hat{z})Var(\xi)$.
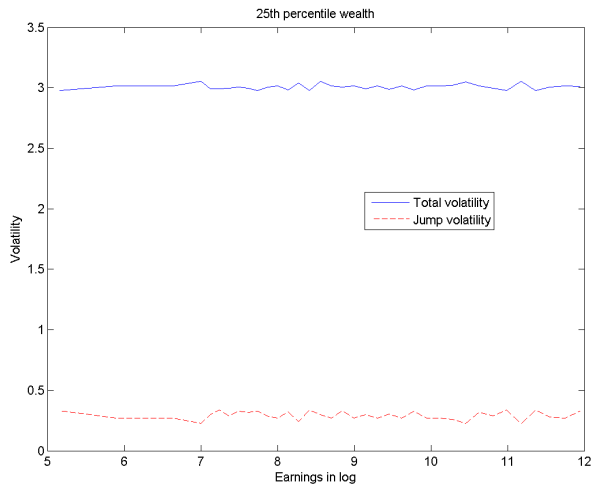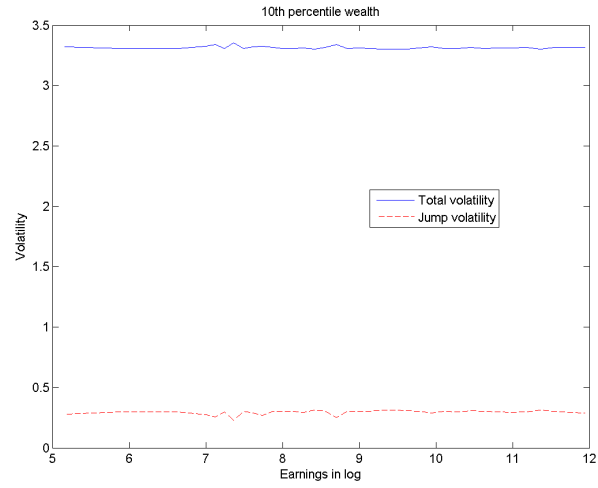
Figure 3.14: Volatility decomposition part 1.
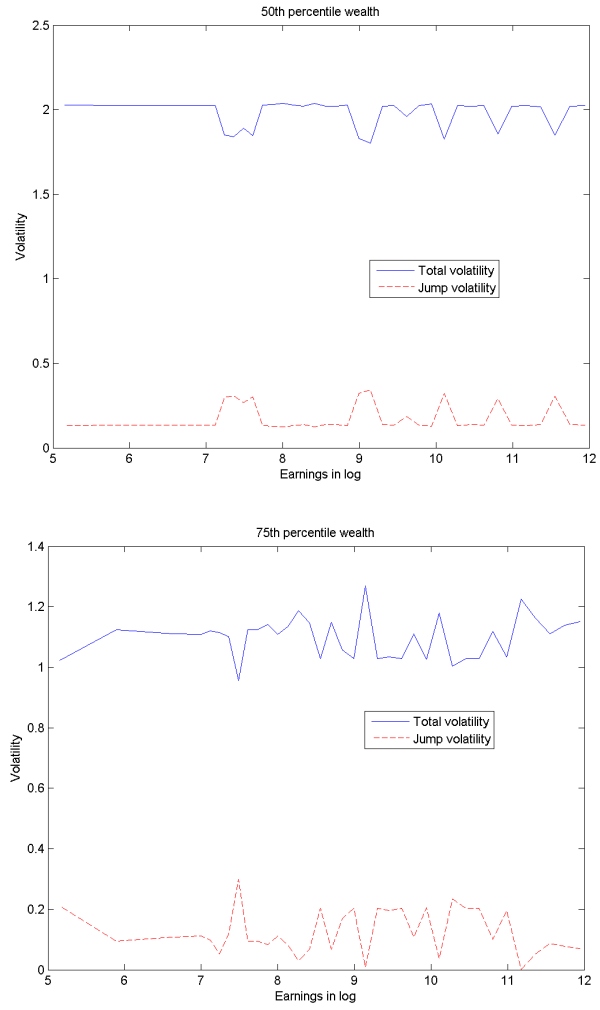
Figure 3.15: Volatility decomposition part 2.
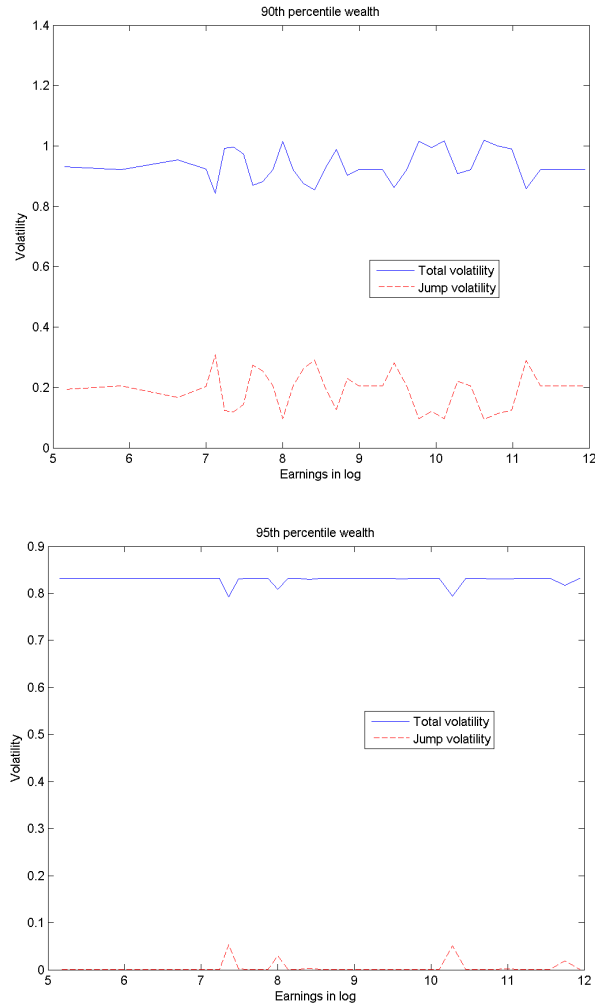
Figure 3.16: Volatility decomposition part 3.

Jump risk accounts for 5% to 15% of the volatility for the less wealthy agents and around 20% of the volatility for the 50-90th percentile wealth agents. For the wealthy agents, the share of the jump risk is relatively small. This suggests that for most agents larger, less frequent deviations in the income process represents an important component in their income process. This type of risk is also harder to hedge. This is in line with the finding in the baseline log normal environment where the less wealthy and wealthy have more consistency between the recovered volatility and the data. If the jump component was misspecified as the variance in the log normal case, then the estimated variance in the log normal case was required to be large to account for these rare large jumps. Then overestimation of the

variances in the log normal case would be expected.

3.5 Utility estimation

The last important element in this simple environment that affects the saving dynamics is the utility function. In this simple environment saving is driven by precautionary motives by agents facing income volatility. Fixing the income process, the level of marginal utility and the degree of risk aversion (the curvature of the utility) function determine the saving behavior. A more risk averse agent would be more sensitive to income volatility by saving more. This translates more income inequality to wealth inequality. I discussed important features of the income process in previous sections. This section will estimate the utility functions that are consistent with the saving behaviors observed.

Traditionally, saving decision could be rationalized through the Euler equation for the H.J.B. equation derived by substituting the F.O.C. condition into the following condition:

$$
\begin{aligned}
(\rho - r)v_{\hat{a}}(\hat{a}, \hat{z}) \;=\; & \partial_{\hat{z}\hat{a}}v(\hat{a}, \hat{z})(\mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma(\hat{a}, \hat{z})^2) \\
& + \partial_{\hat{z}}v(\hat{a}, \hat{z})\frac{\partial(\mu(\hat{a}, \hat{z}) - \frac{1}{2}\sigma(\hat{a}, \hat{z})^2)}{\partial_{\hat{a}}} \\
& + \partial_{\hat{a}\hat{a}}v(\hat{a}, \hat{z})(\frac{s(\hat{a}, \hat{z})}{exp(\hat{a})}) + \frac{1}{2}\partial_{\hat{z}\hat{z}\hat{a}}v(\hat{a}, \hat{z})\sigma(\hat{a}, \hat{z})^2 \\
& + \frac{1}{2}\partial_{\hat{z}\hat{z}}v(\hat{a}, \hat{z})\frac{\sigma(\hat{a}, \hat{z})^2}{\partial_{\hat{a}}},
\end{aligned}
$$

and becomes the Euler equation:

$$
\begin{aligned}
v_{\hat{a}}(\hat{a}, \hat{z}) \;=\; & -E_t \int e^{-(\rho-r)s}[\partial_{\hat{z}}v(\hat{a}_s, \hat{z}_s)\frac{\partial(\mu(\hat{a}_s, \hat{z}_s) + \frac{1}{2}\sigma(\hat{a}_s, \hat{z}_s)^2)}{\partial_{\hat{a}}}]ds \qquad (3.18) \\
& -E_t \int e^{-(\rho-r)s}[\frac{1}{2}\partial_{\hat{z}\hat{z}}v(\hat{a}_s, \hat{z}_s)\frac{\sigma(\hat{a}_s, \hat{z}_s)^2}{\partial_{\hat{a}}}]ds.
\end{aligned}
$$

The marginal utility of an extra amount of consumption would affect the wealth today as well as the future income process. The two forces will balance in the equilibrium.

If we only consider up to two moments then we can solve $v(\hat{a})(\hat{a}, \hat{z})$ for the observed

income process and saving behavior. In the higher moments case, the H.J.B. equation will be different.

$$V(\hat{a}, \hat{z}) = U(c(\hat{a}, \hat{z})) + E \int e^{-\rho s}[V(\hat{a}_s, \hat{z}_s)|\hat{a}, \hat{z}]ds \quad (3.19)$$

The identification becomes more involved. Recently, Schrimpf (2012)(Sch10) discussed several identification requirements. However, the standard method in B.B.L. type two-step estimator still applies if discount factor and interest rate are known. See appendix A.3.2 for more details. We have

$$V(\hat{a}, \hat{z}) = E \int e^{-\rho s}[U(\hat{a}_s, \hat{z}_s)|\hat{a}, \hat{z}]ds \quad (3.20)$$

The equilibrium expectation operator E (transitional matrix) is estimated from the data. The simulation takes following steps:

- First semi-parametrize $U(c(\hat{a}_s, \hat{z}_s))$.

- Simulate sequences of $\hat{a}, \hat{z}$ and calculate $V(\hat{a}, \hat{z})$ according to (3.5). Taking the average will give us $\hat{V}(\hat{a}, \hat{z})$.

- Substitute $\hat{V}$ into the F.O.C. equation

$$U_c(c(\hat{a}, \hat{z})) = -\frac{1}{exp(\hat{a})}\partial_a V(\hat{a}_t, \hat{z}_t)\partial_c s(\hat{a}_t, \hat{z}_t)$$

.

- Find the parameters such that the difference in the F.O.C. is minimized.

To measure the curvature of the utility function, I adopt the standard relative risk aversion measure $-\frac{U''(c)}{U'(c)}c$. I will let $U(.,.)$ satisfy $U(c, \hat{a}) = R(\frac{c}{\hat{a}})\hat{a}^l$. I also impose the shape restrictions $U_c(c, \hat{a}) \geq 0$ and $U_{cc}(c, \hat{a}) \leq 0$ as in Beresteanu (2007)(Ber04). Finally ρ is set to be 0.05 as most previous literature . r is set to be 0.03 to match the 2007-2009 interest rate.

To make the comparison, I also estimate the utility function with a constant relative risk aversion (CRRA) utility function: $U(c) = \frac{c^{1-\gamma}}{1-\gamma}$ for $\gamma \neq 1$ and $ln(c)$ when $\gamma = 1$. γ is the degree of risk aversion. The estimated $\gamma = 0.6254$. This is lower than the range $1-4$ used in

84

many studies. However, it is highly similar to the Hansen singleton's (1983)(Lar82)'s point estimate of 0.68-0.9.[10]

Figure 3.17 plots the estimated results of the degree of relative risk aversion. The non-parametric estimates show strong non-monotonicity. It is lower than the Hansen Singleton estimates except for the case when the consumption wealth ratio is close to 1 and extremely large values(>3). This is in line with the range in Hansen and Singleton's (1983)(Lar82) but lower than most calibration choices.
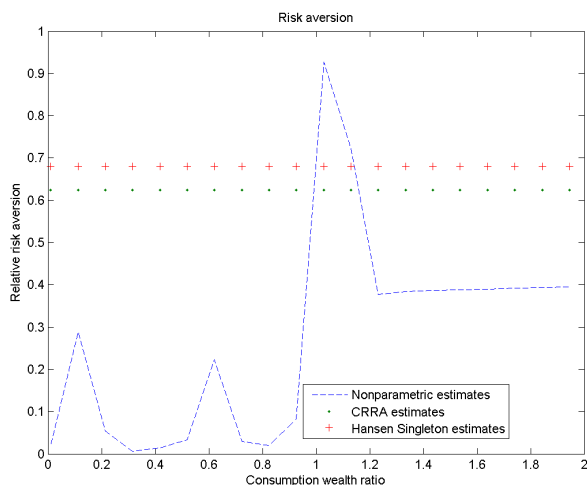
Figure 3.17: Risk aversion

3.6 Risk premium

We have already shown that higher moments are important in explaining saving. How much do these higher moments account for the risk premium? I use the Lucas (1987)(Luc87)'s certainty equivalence definition, which specifies the percentage of the consumption an agent is willing to sacrifice to perfectly smooth the consumption. Following Guvenen et al. (2015)(GKO15):

$$U(c(1 - \pi)) = E(U(c(1 + \delta))),$$

[10]The confidence range is around 0-2 in Hansen and Singleton's (1983)(Lar82) result.

where π is defined as premium, δ is the local shock for the consumption, and σ^2 is the variance of δ. After Taylor expansion we have

$$\pi = -\frac{1}{2}\frac{U''(c,a)c}{U'(c,a)}\sigma^2 - \frac{1}{6}\frac{U^{(3)}(c,a)c^2}{U'(c,a)}skewness*\sigma^3 - \frac{1}{24}\frac{U^{(4)}(c,a)c^3}{U'(c,a)}kurtosis*\sigma^4 \quad (3.21)$$

The last two terms represent the contribution of the higher moment risk. In the constant relative risk aversion case $\frac{U''(c,a)c}{U'(c,a)} = -\gamma, \frac{U'''(c,a)c^2}{U'(c,a)} = (\gamma+1)\gamma$ and $\frac{U'''(c,a)c^3}{U'(c,a)} = -(\gamma+2)(\gamma+1)\gamma$. I also plug in the semi-nonparametric estimates of the utility function. $dlnc \approx d\hat{a}$ if we consider saving rate is roughly constant through perturbation and the conditional variance, skewness, and kurtosis of the percentage change of consumption will be similar to that of the log net wealth $\hat{a}$. The σ^2 will be the local variance of $\hat{a}$.

Figure ??-3.19 presents the estimated results. Solid lines show the risk premium from the second moment. Dash-plus lines stand for the risk premium from the second and the third moment. The dash-dot lines are the risk premium from the second, the third moment and the fourth moment. The blue lines represent the semi-nonparametric utility estimates while green ones represent the CRRA utility function.
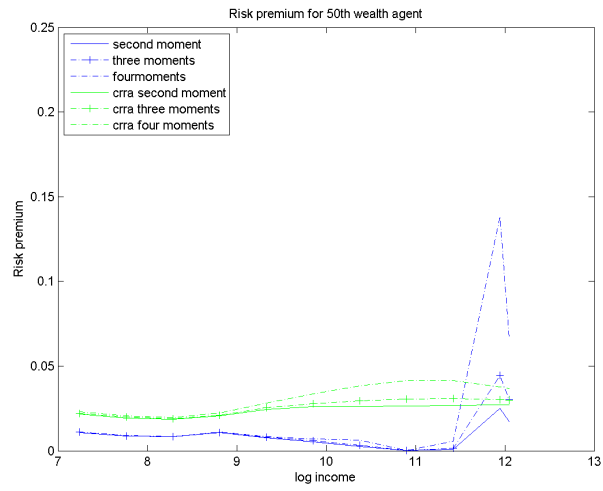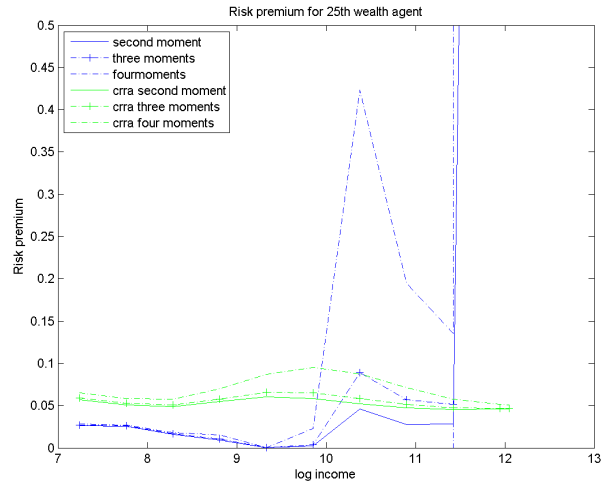
Figure 3.18: Risk premium part 1.
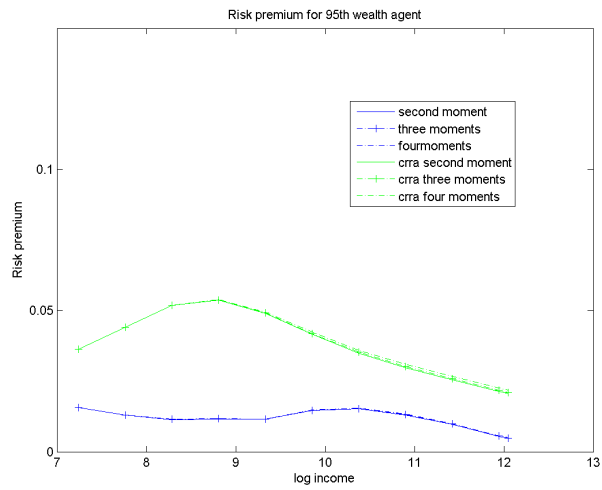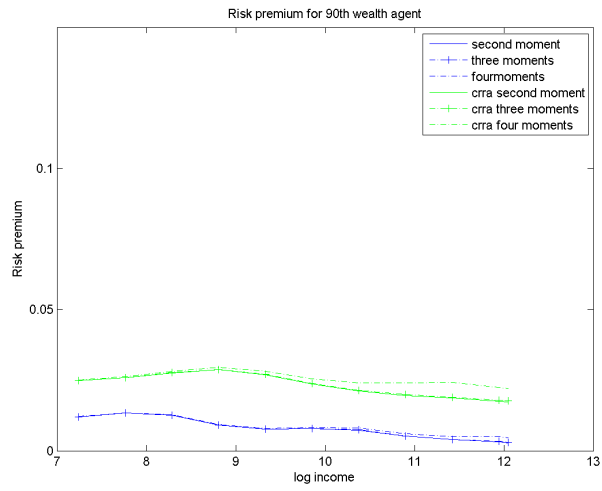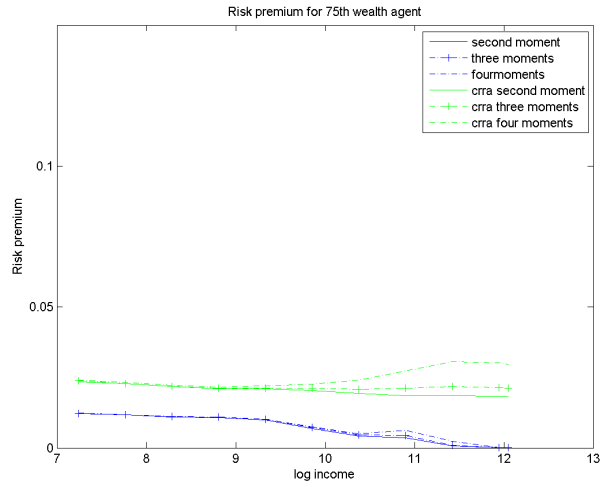
Figure 3.19: Risk premium part 2.

The second moment, which is the traditional risk aversion contributes most to the risk premium. The contribution from the higher moments is highly sensitive to the curvature of the utility function. This could be understood from equation (3.21). $\frac{1}{n!}\sigma^n$ decreases very rapidly. If the curvature term $\frac{U^{(n)}(c)c^{n-1}}{U'(c)}$ increases very rapidly and balances the previous term, then the higher moment is important as a source for the risk premium. The risk premium coming from the third moment is around 0.01% to 2%,contribution from the forth moment is around 0.2% to 0.5%[11] in the nonparametric case .In the CRRA case, the degree of risk aversion is higher than the nonparametric estimate, resulting a higher risk premium and a larger effect. 0.1 % to 0.6% from the third moment and 0.2 % to 2.5% from the forth moment. Since $\hat{a}$ and $\hat{z}$ are correlated, the higher-order moments of log income process contributes to the risk premiums. This contribution is larger for the less wealthy agent, which confirms the earlier finding that non-normality feature is stronger for the less wealthy agent.

In Guvenen et al. (2015)(GKO15) σ is fixed to be 0.1 and skewness is -2 to 0 , kurtosis is 3 to 30, which is from the income process. The utility function is chosen to be CRRA with degree of risk aversion of 10. Their result implies 400% increase in risk premium by incorporating the higher moments. This large increment is very sensitive to the high risk aversion. If the risk aversion is 0.6245 as in the SCF data, this increment reduces to 23%.

3.7   Conclusion

Income wealth dynamics is important in understanding inequalities in society. This paper addressed the basic forces that affects the saving choice, which in turn influences the income wealth joint distribution, namely the income process and marginal utility. I showed that higher moments of the income process are important in explaining cross-sectional saving behavior and income wealth joint distribution. Higher moments also provide extra source for risk premium.

To illustrate the main forces, production, firm ownership, and financial market were

---

[11]except for some extreme region the premium is around 10 %

excluded in this simple environment. It would be interesting to incorporate these factors and examine their effects with richer data and a richer framework in future.

APPENDIX A

Appendix

A.1   Appendix for chapter 1

1	2	3	4	5	6
cathode	seq	cantilever	fluid	wire	recording
emission	id	tip	liquid	wires	medium
anode	sequence	force	chamber	nanoscale	information
emitter	amino	end	flow	crossbar	reproducing
display	protein	deflection	membrane	array	apparatus
field	polypeptide	free	inlet	molecular	recorded
electron	encoding	atomic	outlet	junction	probe
electrons	expression	probe	reservoir	set	means
emitting	vector	microscope	communication	junctions	detecting
phosphor	acid	arm	channel	switch	probes

11	12	13	14	15	16
magnetic	storage	solution	barrier	light	catalyst
magnetization	data	aqueous	layers	optical	catalytic
ferromagnetic	writing	solvent	band	wavelength	growing
field	storing	salt	energy	waveguide	growth
nonmagnetic	domain	water	carriers	near-field	grown
magnetoresistive	reading	acid	semiconductor	emitting	vapor
magnetically	stored	adding	layer	aperture	deposition
non-magnetic	information	mixture	gap	incident	grow
spin	read	surfactant	structure	reflected	iron
current	media	organic	electrons	optically	support

Figure A.1: Keywords of the topics, part 1.

7	8	9	10
nanocrystal	gate	determining	composite
nanocrystals	drain	measuring	matrix
core	transistor	measuremen	fibers
shell	source	measured	fiber
semiconductor	channel	data	nanofibers
zns	region	determined	ceramic
cds	effect	step	nanocomposite
outer	dielectric	determine	polymer
population	semiconductor	function	material
iii-v	field	calculating	resin

17	18	19	20
qubit	particles	signal	cnt
superconducting	particle	frequency	cnts
junction	size	output	array
josephson	nm	amplitude	carbon
qubits	average	means	alignment
quantum	diameter	detecting	functionalized
state	fine	generating	comprising:a
loop	mu.m	feedback	template
computing	colloidal	detection	nanotube
mesoscopic	silica	vibration	arrays

Figure A.2: Keywords of the topics, part 2.

21	22	23	24	25	26
memory	member	element	titanium	probe	nanotubes
cell	movement	elements	copper	scanning	carbon
line	support	rare	aluminum	microscope	single-wall
cells	mounted	earth	nickel	sample	single-walled
bit	stage	ribbon	iron	tip	metallic
word	movable	structure	oxide	tunneling	semiconducting
lines	actuator	electromechanical	cobalt	means	swnts
resistance	piezoelectric	arranged	zinc	image	combinations
write	mechanism	switching	silver	scan	swnt
access	fixed	piezoelectric	metal	distance	functionalized

31	32	33	34	35	36
beam	gas	nanoparticles	nanowire	fullerene	composition
laser	reactor	nanoparticle	nanowires	fullerenes	weight
electron	reaction	metal	semiconductor	c.sub.60	total
focused	chamber	nm	growth	derivative	amount
ion	inert	gold	grown	molecules	percent
beams	pressure	precursor	p-type	atoms	oil
energy	plasma	size	growing	carbon	mixtures
directing	zone	inorganic	n-type	molecule	compositions
focusing	gaseous	silver	length	solvent	resin
source	hydrogen	organic	doped	molecular	wt

Figure A.3: Keywords of the topics, part 3.

27	28	29	30
pattern	quantum	polymer	temperature
mold	dot	copolymer	precursor
patterned	dots	monomer	heating
resist	emitting	block	powder
template	nm	polymeric	heated
patterning	barrier	polymers	heat
imprint	semiconductor	copolymers	mixture
substrate	active	monomers	atmosphere
lithography	zns	groups	hours
patterns	laser	polyethylene	period

37	38	39	40
nanotube	switching	ni	film
carbon	input	fe	thin
nanotubes	output	cu	films
structure	logic	pt	substrate
single-walled	terminal	pd	manufacturing
end	node	au	thickness
multi-walled	state	ti	crystal
length	terminals	mo	organic
diameter	channel	cr	formed
outer	communication	ag	silicon

Figure A.4: Keywords of the topics, part 4.

41	42	43	44	45	46
alkyl	agent	circuit	nucleic	etching	recited
formula	administering	integrated	molecule	layer	assembly
substituted	drug	conductor	binding	depositing	thermal
compound	delivery	chip	target	silicon	deposited
atoms	therapeutic	electronic	molecules	forming	diamond
independently	agents	device	bound	dielectric	substantially
aryl	effective	devices	dna	mask	disposed
atom	pharmaceutical	plurality	acid	removing	pores
hydrogen	pharmaceutically	conductors	complex	substrate	thermally
groups	administered	dielectric	oligonucleot	deposition	making

Figure A.5: Keywords of the topics, part 5.

47		48	49	50
nanostructure	direction	electrode	conductive	
nanostructures	parallel	electrodes	electrically	
one-dimensional	extending	voltage	insulating	
substrate	substantially	current	electrical	
nanorod	perpendicular	applying	contact	
nanorods	axis	electric	layer	
elongated	width	connected	conducting	
methods	longitudinal	applied	conductivity	
disposed	plane	electrolyte	material	
growth	length	potential	sensor	

Figure A.6: Keywords of the topics, part 6.

A.2　Appendix for chapter 2

A.2.1　Properties of the equilibrium

It is worth mentioning that among these papers the equilibrium distribution of the heterogeneity usually has a fat tail. The origin and properties of fat tail distribution has been a important topic in economics, network, computer science, and many other subjects. We show the conditions when the equilibrium distribution will have a fat tail.

Corollary 1. *If $\bar{\mu}(z) = bz$ then the cdf becomes $\frac{1}{1+\frac{1}{C}\left(\frac{z}{z}\right)^{-\frac{\alpha}{\lambda-b}}}$ which is the log logistic distribution, the solution to (LM14).*

Remark Logistic type distribution and diffusion process has been used frequently in empirical technology spill over studies. Apart from this popular class, when $\widetilde{\mu(x)} \sim a(\frac{x-m}{s})^{1+a}$ the density $\tilde{F}(z)$ approaches *Fréchet* distribution. These two classes of distributions were widely studied in finance, macro and international trade literature. (Gab09) surveyed related applications of fat tail distribution in economics and finance.

A.2.2　Law of motion equation

This third term is approximated in the following sense: in each time period the inflow due to the drifting agents is $+\int h(\delta, x-\delta, t)f(x-\delta, t)d\delta$ and the outflow is $-\int h(\delta, x, t)f(x, t)d\delta$. $h(\delta, x, t)$ is some instantaneous rate function of moving from x to δ away at time t. Combining the two we have $\int[h(\delta, x-\delta, t)f(x-\delta, t) - h(\delta, x, t)f(x, t)]d\delta$ and Taylor expand the first term with respect to $x-\delta$ at x in the integrand up to second order and exchange differentiation and integration we have $\int[h(\delta, x, t)f(x, t) - \delta\frac{\partial}{\partial x}(h(\delta, x, t)f(x, t)) + \frac{\delta^2}{2}\frac{\partial^2}{\partial x^2}(h(\delta, x, t)f(x, t)) - h(\delta, x, t)f(x, t)]d\delta = -\frac{\partial}{\partial x}(\int \delta h(\delta, x, t)f(x, t)d\delta) + \frac{\partial}{\partial x^2}(\int \frac{\delta^2}{2}h(\delta, x, t)f(x, t)d\delta)$.

$\int \delta h(\delta, x, t)d\delta = \mu(x, t)$ and $\int \frac{\delta^2}{2}h(\delta, x, t)d\delta = 0$

A.2.3 Proof of proposition I

Proof: Let us denote $\Phi(z) = 1 - \tilde{F}(x)$ and slightly abuse the notation and let $\bar{\mu}$ stands for the equilibrium level of R&D.

The equation (2.5) in equilibrium takes the form:

$$\lambda z \phi(z) = -\alpha \Phi(z)(1 - \Phi(z)) + \bar{\mu}(z)\phi(z) \tag{A.1}$$

The evolution could be solved as follows

$$\phi(z) = \frac{\alpha}{\bar{\mu}(z) - \lambda z} \Phi(z) - \frac{\alpha}{\bar{\mu}(z) - \lambda z} \Phi(z)^2$$

Denote $\frac{\alpha}{\bar{\mu}(z) - \lambda z} = P(z)$ Then the equation becomes $\phi(z) \equiv P(z)\Phi(z) - P(z)\Phi(z)^2$.

Let $\Phi(z) = \frac{1}{y(z)}$ we would have $y(z) = [e^{\int_{\underline{z}}^{z} P(s)ds}]^{-1}(C + \int_{\underline{z}}^{z} P(s)e^{\int_{\underline{z}}^{s} P(s)ds}ds)$.

Then

$$\begin{aligned}
\Phi(z) \quad &= [e^{\int_{\underline{z}}^{z} P(s)ds}][C + \int_{\underline{z}}^{z} P(s)e^{\int_{\underline{z}}^{s} P(\psi)d\psi}ds]^{-1} \\
&= \frac{e^{\int_{\underline{z}}^{z} P(s)ds}}{C + e^{\int_{\underline{z}}^{z} P(s)ds}}
\end{aligned}$$

We need to check $\tilde{F}(z)$ is indeed a distribution function. Since $P(s) < 0$, then $e^{\int_{\underline{z}}^{z} P(s)ds - lnC}$ is in $(0,1)$, and $\tilde{F}(z)$ is between $(0,1)$. When z goes to infinity $M(z)$ goes to negative infinity by condition $\widetilde{\mu(.)} \sim x^d$ and $d \leq 1$. $lim_{z \to \infty}\tilde{F}(z) \to 1$

We next check the monotonicity. By assumption: $\partial_z M(z) = P(z) < 0$, $\tilde{F}(z)$ is increasing with z.

Remark There is a degenerate mass at the lower bound of size $\frac{C}{1+C}$. In (BPT14) there is a similar result. In their paper, imitation is strong enough that relative difference shrinks for the lower technology owner while the technology leader keep their growth rate through R&D. The population with sufficient high technology level preserves the initial dispersion. In our environment this degenerate mass was determined by initial condition.

A.2.4 Proof of lemma 1

We can rewrite (2.5) and denote $\Phi(z) = 1 - \tilde{F}(x)$ as the tail probability. Slightly abuse the notation and let $\bar{\mu}(z)$ stands for the equilibrium level of R&D. Further divide both side

by $\Phi(z)$:

$$\lambda \frac{z\Phi'(z)}{\Phi(z)} = -\alpha(1 - \Phi(z)) + \frac{\bar{\mu}(z)}{z} \frac{z\Phi'(z)}{\Phi(z)}$$

This should hold in the limit as z goes to infinity then we have

$$\lambda \lim_{z \to \infty} \frac{z\Phi'(z)}{\Phi(z)} = -\alpha + \lim_{z \to \infty} \frac{\bar{\mu}(z)}{z} \frac{z\Phi'(z)}{\Phi(z)} \tag{A.2}$$

A.2.5 A second empirical approach

In this case, we assume only meeting leaves a trace, self R&D happened after the meeting. The transitional probability $T_i(.|.)$ is not time invariant. When we consider the invariant distribution along the balance growth path this will become $T(ae^{-\lambda(i+1)}|be^{-\lambda i})$, and $T(.|.)$ would be time invariant. Particularly

$$T(ae^{-\lambda(i+1)}|be^{-\lambda i}) = \alpha \mathbf{I}_{a>b} \tilde{f}_i(a) = \alpha \mathbf{I}_{a>b} e^{-\lambda(i+1)} \tilde{f}_1(ae^{-\lambda(i+1)})$$

, $\tilde{f}_1(.)$ is the invariant distribution. This result follows the assumption that patent only records the meeting process. In a random meeting environment, the probability of meeting a target with a specific type is the probability share of the target type multiply by the meet intensity α which becomes $\alpha \tilde{f}_i(a)$

We will assume any potential upgrade or downgrade in the R&D process $\mu(.)$ will happen after the meeting. Then the likelihood function could be rewritten as

$$f(y_1, y_2) = \int_{z_1} \int_{z_2 > z_1} p_2(y_2|z_2) e^{-\lambda(2)} \tilde{f}_1(z_2 e^{-\lambda(1)}) p_1(y_1|z_1) \tilde{f}_1(z_1) dz_1 dz_2 \tag{A.3}$$

A.2.6 Identification of Poisson Mixture

The following proof was adopted from (Rao92) p214. When we have t=0, the mixture takes the form of

$$H(y) = \int P(y, m) \Gamma(dm)$$

$H(.)$ is the cdf of y, $\Gamma(.)$ is the cdf of the possible poisson rates Then the characteristic function satisfies

$$\phi_H(t) = \int \phi_P(t, m) \Gamma(dm)$$

since poisson distribution belongs to the additive group so $\phi_P(t, m) = \phi_P(t, 1)^m$

Let $\psi_\Gamma(l) = \int l^k d\Gamma(k)$. For any two candidates $F_1(.), F_2(.)$ we have $\psi_{\Gamma_1}(l) = \psi_{\Gamma_2}(l)$ furthermore replace l with ρe^{it}. In particular it holds for $l = \rho e^{it}$ where $\rho \in (0, 1)$ Applying the dominated convergence theorem, which holds for $\rho = 1$

We have for any two distributions

$$\int e^{itk} d\Gamma_1(m) = \int e^{itk} d\Gamma_2(m)$$

for any t,which suggests that Γ is identified by uniqueness of Laplace transformation.

There are two circumstances when we discretize m and calculate $\Gamma(dm)$

1. support of m is known

2. support of m is unknown

When support of m is known we need to calculate the $\Gamma(dm)$ through EM algorithm or NPMLE. $\Gamma(dm)$ is between $0, 1$ when they are in discrete form, the domain is then convex. The likelihood attains a maximum in its convex hull.

When we don't know the support of m, according to (Lin95), there are half of the elements in m we could identify comparing to the first case. Then we can perform the same procedure. A heuristic understanding is that both $\Gamma(m_i)$ and m_i are unknowns, there are N moments equations from the data and one regularity condition $\sum \Gamma(dm_i) = 1$. There are $2*\#\{m\}$ unknowns and N+1 equations as long as $\#\{m\}$ is less or equal to $\frac{N+1}{2}$ we can identify the support and its weight.

Other periods apply this same idea similarly.[1]

A.2.7   The joint density of poisson rates m_t are identified

The proof is adopted from (Tei67). If the marginals are identified, then the joint are identified. It suffices to show the joint distribution with one more dimension is identified. In previous section we showed that poisson mixture is identified.

---

[1]This is similar to multivariate structure. When marginals are identified the joint are identified.

Denote $P(x; \alpha)$ as the *c.d.f* for poisson distribution of rate α, $P(x; \alpha)$ be the *c.d.f* for poisson distribution of rate β, and $G(.,.)$ be the joint distribution. $G_2()$ denotes the marginal distribution. $P(x; \alpha)$ is identified. We need to show $G(.,.)$ is identified.

$$\int_{\alpha,\beta} P(x; \alpha)P(y; \beta)dG(\alpha, \beta) = \int_{\alpha,\beta} P(x; \alpha)P(y; \beta)d\hat{G}(\alpha, \beta)$$

This could be rewritten as [2]

$$\int_{\beta} P(y; \alpha)H(x; \beta)dG_2(\beta) \quad = \int_{\beta} P(y; \beta)\hat{H}(x; \beta)d\hat{G}_2(\beta), \tag{A.4}$$

$$H(x; \beta) \quad = \int_{\alpha} P(y; \alpha)dG(\alpha|\beta), \tag{A.5}$$

$$\hat{H}(x; \beta) \quad = \int_{\alpha} P(y; \alpha)d\hat{G}(\alpha|\beta). \tag{A.6}$$

This could be re-expressed as

$$\int_{\beta} P(y; \alpha)dJ(\beta) \quad = \int_{\beta} P(y; \alpha)d\hat{J}(\beta), \tag{A.7}$$

$$J(\beta) \quad = \int_0^{\beta} H(x; \gamma)dG_2(\gamma), \tag{A.8}$$

$$\hat{J}(\beta) \quad = \int_0^{\beta} \hat{H}(x; \gamma)d\hat{G}_2(\gamma). $$

Since the Poisson mixture is identified in previous section, we have

$$\int_0^{\beta} H(x; \gamma)dG_2(\gamma) = \int_0^{\beta} \hat{H}(x; \gamma)d\hat{G}_2(\gamma) \tag{A.9}$$

and let $x \to \infty$ we have from equation (A.4) that

$$\int_{\beta} P(y; \alpha)dG_2(\beta) = \int_{\beta} P(y; \beta)d\hat{G}_2(\beta).$$

Applying the identifiability of the Poisson mixture model again, we have

$$G_2(.) = \hat{G}_2(.)$$

Then from equation (A.9)

$$H(x; \beta) = \hat{H}(x; \beta)$$

---

[2]This requires the joint density could be expressed in the product space, see (Nev65) page 75 for details

102

then this says

$$G(.|\beta) = \hat{G}(.|\beta)$$

Then the joint distribution

$$G(.|.) = \hat{G}(.|.)$$

A.2.8 Assumption of patent distribution

The assumptions that the distribution of the patents follows $\tilde{f}(z)$ is an extra assumption we used for simplicity. In (Mat03), as long as we have identified $\tilde{F}(ze^{-\lambda t})$ with normalizations, both the distribution of z and function $\tilde{F}(.)$ could be identified. By choosing the right normalization and fix $\hat{t}$ then the observed density of $\tilde{F}(;\hat{t})$ is the density of z. Then substitute back the identified density, the functional of $\tilde{F}(.)$ could be identified.

A.2.9 Extension when there is constant Brownian motion

Bellman equation (2.1) becomes

$$(\rho - \lambda)v(z) = \max_{\bar{\mu}(z)} \bar{\pi}(z) - \bar{\Psi}(\bar{\mu}(z)) + (\bar{\mu}(z) - \lambda z)v'(z) + \frac{\sigma^2}{2}v''(z) \tag{A.10}$$

$$+\alpha \int_z [v(z') - v(z)]\tilde{f}(z')dz'$$

Integrate the law of motion equation (2.2) with respect to x and rescale we have the law of motion equation of the cumulative distribution. It is a third order nonlinear ordinary differential equation of $\tilde{F}(z)$:

$$\alpha\tilde{F}(z)(1 - \tilde{F}(z)) = (\lambda z - \bar{\mu}(z))\tilde{f}(z) + \frac{\sigma^2}{2}\partial_z\tilde{f}(z) \tag{A.11}$$

$$\bar{\mu}(z) = \lambda z - \frac{\alpha\tilde{F}(z)(1 - \tilde{F}(z))}{\tilde{f}(z)} + \frac{\sigma^2}{2}\frac{\partial_z\tilde{f}(z)}{\tilde{f}(z)} \tag{A.12}$$

If σ is identified in some pre-step, the same proof logic applies.

A.2.10 The original Lucas and Moll (2014)(LM14)

The law of motion equation is the following:

$$-\phi(x)\gamma - \phi'(x)\gamma x = \phi(x)\int_0^x \alpha(y)\phi(y)dy - \alpha(x)\phi(x)[1-\Phi(x)]. \qquad (A.13)$$

Given $\phi(x)$ this is an integral equation of $\alpha(x)$ we can perform the standard inversion operator K^* that $K^*\alpha(x) = \int_0^x \alpha(y)\phi(y)dy$

$$-\phi(x)\gamma - \phi'(x)\gamma x = \phi(x)[K^*(x) - (1-\Phi(x))]\alpha(x)$$

and

$$\alpha(x) = \qquad [[\phi(x)K^* - \phi(x)(1-\Phi(x))]' * [\phi(x)K^* - \phi(x)(1-\Phi(x))]]^{-1}$$
$$[\phi(x)K^* - \phi(x)(1-\Phi(x))]\gamma(-\phi(x) - \phi'(x)x)$$

Specifically we express $\alpha(x)$ with orthornomal basis function $h_k(x)$. $\alpha(x) = \frac{1}{1+\sum_k^\infty [\beta_k h_k(x)]^2}$. For given $\phi(x)$ and γ we can find β_ks that minimize the law of motion equation.

Then we can go back to the first step. The inverse optimal control equation is

$$(\rho + \alpha(x)(1-F(y)))v'(x) = 1 - \alpha^{-1}(x) - \gamma x v''(x). \qquad (A.14)$$

If $\alpha(x)$ is specified and is monotone, then we can apply the same method to recover $v'(x)$.

A.2.11 The original Luttmer (2007)(Lut07) model with choice of drift

To fix the idea consider the following example based on Luttmer (2007)(Lut07) firm dynamic model:

Rewrite everything denominated by fixed cost, denote s: firm size relative to fixed costs(Luttmer (2007) (Lut07) for details). With some assumptions it follows a brownian motion. a is age, $\pi(\mu(s))$ is the flow profit. Firm can choose their optimal drift $\mu(s)$ costly:

$$ds = \mu(s)da + \sigma dW_{t/a} \qquad (A.15)$$

104

where $W_{t/a}$ is a white noise. We focus on a stationary environment and assume σ is constant.

Value of a firm would be

$$V(s) = \max_{\tau} E[\int_0^{\tau} e^{-ra}\pi(\mu(s))da] \tag{A.16}$$

It has a cutoff rule , if $s > b$ firm remains open and the Hamiltonin-Jacobi-Bellman equation is

$$rV(s) = \pi(\mu(s)) + V'(s)\mu(s) + \frac{1}{2}V''(s)\sigma^2 \tag{A.17}$$

With value matching and smooth pasting conditions $V(b) = 0, V'(b) = 0$. The measure of firms of age a and size s at time t satisfies $f(a, s, t) = m(a, s)$ and $da = dt$. The law of motion in the economy follows Kolmogorov forward equation

$$\frac{\partial m(a, s)}{\partial a} = -\frac{\partial}{\partial s}[\mu(s)m(a, s)] + \frac{1}{2}\frac{\partial^2}{\partial s^2}[\sigma^2 m(a, s)] \tag{A.18}$$

Equation (A.17), Equation (A.18) and the boundary conditions describe the economy.[3]

First step: $m(a, s)$ has an empirical counter part, techniques similar to Aït-Sahalia (1996)(Ait96) could be applied, integrate equation (A.18) to s :

$$\int^s \frac{\partial m(a, l)}{\partial a}dl + \mu(s)m(a, s) = \frac{1}{2}[\sigma^2 m(a, s)]'$$

If the Brownian term σ is estimated in separately, we can reconstruct the drift $\mu(s)$ from this equation. The identification comes from the uniqueness of the reconstruction.

Second step: plug the equilibrium $\mu(s)$ and σ into (A.17). $V(s)$ and $\pi(\mu(s))$ are reconstructed through inverse optimal control problem. We have $rV'(s) = V''(s)\mu(s) + \frac{1}{2}V'''(s)\sigma^2$.

In an alternative parametric setup we could use forward simulations to simulate $V(s, \theta) = E[\int_0^{\tau} e^{-ra}\pi(s, \theta)da]$ given θ. Then find θ through minimum distance estimator: $\theta_0 = argmin_{\theta}|(r)V(s, \theta) - [\pi(s, \theta) + V'(s, \theta)\mu(s) + \frac{1}{2}V''(s, \theta)\sigma^2]|$. This could be jointly estimated with the law of motion equation in one step GMM.

---

[3] In this simple environment H.J.B. didn't depend on the macro distribution.

A.2.12 Recovering value function

We have

$$\rho v(x) \quad = \bar{\pi}(x) - \bar{\Psi}(\bar{\mu}(x)) + \bar{\mu}(x)v'(x) + \lambda v(x) - \lambda x v'(x) + \alpha S(x) \qquad (A.19)$$

$$S(x) \quad = \int_x^\infty [v(y) - v(x)] \, f(y) dy \qquad (A.20)$$

$$F(x) \quad = \frac{1}{1 + e^{-\int_z^x \frac{\alpha}{\lambda s - \mu(s)} ds - lnC}} \qquad (A.21)$$

Notice we empirically recovered $\mu(x)$ and $F(x)$ The optimality condition requires

$$\bar{\Psi}'(\bar{\mu}(x)) = v'(x) \qquad (A.22)$$

which leads to the result

$$\bar{\Psi}(\bar{\mu}(x)) = \int v'(x)\bar{\mu}'(x)dx = v'(x)\bar{\mu}(x) - \int \bar{\mu}(x)dv'(x)$$

up to a scaler shift.

If there is information on the value function like market value through stock market, $\bar{\Psi}(.)$ could be inferred from here.

If $v(.)$ is unknown, with a specification of $\bar{\pi}(.)$ and substitute this back into equation (A.19) we have

$$(\rho - \lambda)v(x) \quad = \bar{\pi}(x) + \int \bar{\mu}(x)dv'(x) - \lambda x v'(x) + \alpha S(x) \qquad (A.23)$$

With a specification of $\bar{\pi}(.)$ and empirical knowledge $F(.), \bar{\pi}(x)$ we can calculate $v(.)$ by combining (A.23) and (A.20)

$$(\rho - \lambda)v(x) = \quad \bar{\pi}(x) + \sum \bar{\mu}(x)v''(x)\Delta x - \lambda x v'(x)$$
$$+ \alpha \left[\sum v(i)f(i)\Delta i - v(x)(1 - F(x)) \right].$$

Now define the following:

$$v'(x) \quad \approx \frac{v_j - v_{j-1}}{\Delta x}$$

$$v''(x) \quad \approx \frac{v_{j+1} - 2v_j + v_{j-1}}{(\Delta x)^2}$$

rearrange equation we have

$$(\rho - \lambda)v_j = \quad \bar{\pi}(x) + \sum_{i=1}^{j} \bar{\mu}_i \frac{v_{i+1} - 2v_i + v_{i-1}}{\Delta x} - \lambda x \frac{v_j - v_{j-1}}{\Delta x}$$

$$+\alpha \left[\sum_{i=j}^{I} v(i)f(i)\Delta i - v_j(1 - F(x)) \right],$$

which is

$$(\rho - \lambda + \frac{\lambda x}{\Delta x} + \alpha(1 - F(x)))v_j - \frac{\lambda x}{\Delta x}v_{j-1}$$

$$-\sum_{i=1}^{j} \frac{\bar{\mu}_i}{\Delta x}v_{i+1} + 2\sum_{i=1}^{j} \frac{\bar{\mu}_i}{\Delta x}v_i - \sum_{i=1}^{j} \frac{\bar{\mu}_i}{\Delta x}v_{i-1} - \alpha \sum_{i=j}^{I} v(i)f(i)\Delta x = \pi(x)$$

$Av^{n+1} = b^n$, $A = B - E + 2F - G - C$ and $b = \pi(x)$ we are going to iterate and use

$$\frac{1}{\Delta}(v^{n+1} - v^n) + Av^{n+1} = b$$

Since the specification of $\pi(.)$ is fixed and μ, F are empirically estimated, we should have $Av = b$ and $v = A^{-1}b$.

For inference purposes, $var(v)$ could be inferred from continuous mapping theorem.

To see this linearity structure we can differentiate equation(A.19) and

$$[\rho + \alpha(1 - F(x))] v'(x) = \pi'(x) + [\bar{\mu}(x) - \lambda x] v''(x) \tag{A.24}$$

This is a first order ordinary equation. One can use a two step parametric estimator to provide a confidence interval.

A.3 Appendix for chapter 3

A.3.1 Comparison

In this section I compare the estimates of conditional skewness and kurtosis to the Guvenen et al. (GKO15)'s result.
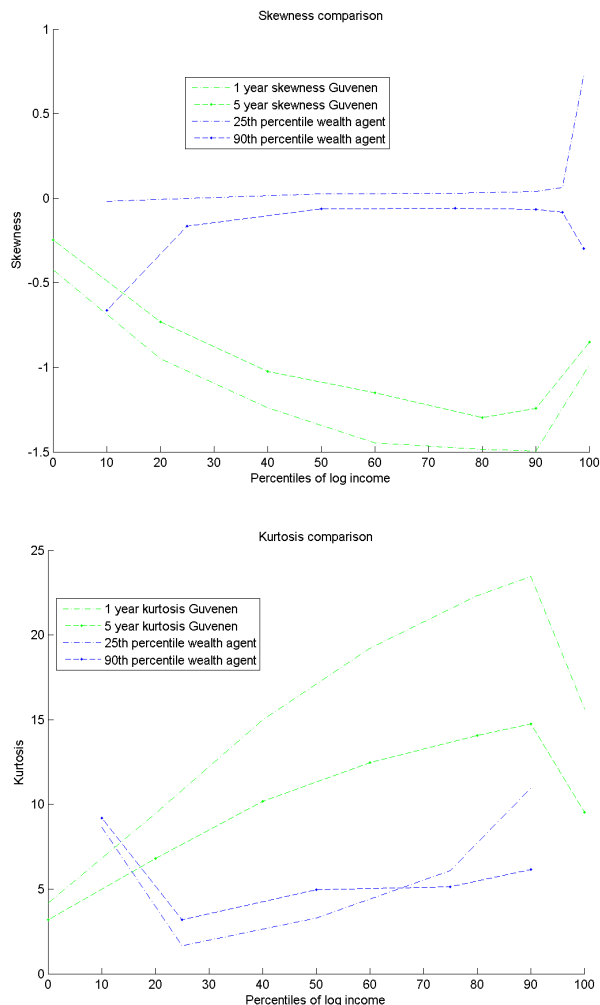
Figure A.7: Conditional skewness and kurtosis

The conditional skewness is smaller than Guvenen et al. (2015)(GKO15)'s estimates and show a hump shape than U shape. The conditional kurtosis is also smaller than Guvenen et al. (2015)(GKO15)'s estimates but show a similar shape. Since SCF 07-09 is a shorter panel and experiences the financial crisis, this sub sample might show different features.

A.3.2   Recovering utility function in higher-order moments case

First order approach:

In the section 3.4.1 we introduced the so called Kramer-Moyal expansion of the transi-

tional matrix for the law of motion equation. Similarly we have the Kramer-Moyal backward expansion of the transitional matrix in the utility function.

$$V(\hat{a}, \hat{z}) = U(c(\hat{a}, \hat{z})) + \sum \beta M_n(m|\hat{a}, \hat{z}) \frac{1}{n!} \frac{\partial^n}{\partial m^n} V(\hat{a}, \hat{z})$$

.

Since consumption choice only affect the first moment of $\hat{a}$ so the first order condition becomes $u'(c_t) = -\frac{1}{exp(\hat{a})} \partial_a v(\hat{a}_t, \hat{z}_t) \partial_c s(\hat{a}_t, \hat{z}_t)$.

Bibliography

[AAP14] David Abrams, Ufuk Akcigit, and Jillian Popadak. "Patent Value and Citations: Creative Destruction or Strategic Disruption." *NBER working paper*, 2014.

[ABL08] Fernando E. Alvarez, Francisco J. Buera, and Robert E. Lucas. "Models of Idea Flows." *NBER working paper*, 2008.

[ABL14] Y Achdou, Francisco J. Buera, J-M Lasry, P-L Lions, and Benjamin Moll. "Partial differntial equation models in macroeconomics." *Phil Trans. R Soc A*, **372**, 2014.

[ABR13] Fernando E. Alvarez, Francisco J. Buera, and Jr. Robert E. Lucas. "Idea Flows, Economic Growth, and Trade." NBER Working Papers 19667, National Bureau of Economic Research, Inc, November 2013.

[ACG14] Ufuk Akcigit, Murat Alp Celik, and Jeremy Greenwood. "Buy, Keep or Sell: Economic Growth and the Market for Ideas." (13-069), December 2014.

[ACH12] Daniel Ackerberg, Xiaohong Chen, and Jinyong Hahn. "A Practical Asymptotic Variance Estimator for Two-Step Semiparametric Estimators." *The Review of Economics and Statistics*, **94**(2):481–498, May 2012.

[Ait96] Yacine Aït-Sahalia. "Nonparametric Pricing of Interest Rate Derivative Securities." *Econometrica*, **64**(3):527–60, May 1996.

[Aiy94] S Rao Aiyagari. "Uninsured Idiosyncratic Risk and Aggregate Saving." *The Quarterly Journal of Economics*, **109**(3):659–84, August 1994.

[B06] Fernando Leiva B. "Pricing Patents through Citations." 2006 Meeting Papers 834, Society for Economic Dynamics, December 2006.

[BBL07] Patrick Bajari, C. Lanier Benkard, and Jonathan Levin. "Estimating Dynamic Models of Imperfect Competition." *Econometrica*, **75**(5):1331–1370, 09 2007.

[Bel70] Richard Bellman. "Dynamic programming and inverse optimal problems in mathematical economics." *Journal of Mathematical Analysis and Applications*, **29**(2):424 – 428, 1970.

[Ber04] Arie Beresteanu. "Nonparametric Estimation of Regression Functions under Restrictions on Partieal Derivatives." Technical report, 2004.

[Bes08] James Bessen. "The value of U.S. patents by owner and patent characteristics." *Research Policy*, **37**(5):932–945, June 2008.

[Bew77] Truman Bewley. "The permanent income hypothesis: A theoretical formulation." *Journal of Economic Theory*, **16**(2):252–292, December 1977.

[BN03] Federico M. Bandi and Thong H. Nguyen. "On the functional estimation of jump-diffusion models." *Journal of Econometrics*, **116**(12):293 – 328, 2003. Frontiers of financial econometrics and financial engineering.

[BP03] Federico M. Bandi and Peter C. B. Phillips. "Fully Nonparametric Estimation of Scalar Diffusion Models." *Econometrica*, **71**(1):241–283, 2003.

[BPT14] Jess Benhabib, Jesse Perla, and Christopher Tonetti. "Catch-up and fall-back through innovation and imitation." *Journal of Economic Growth, Springer*, **19**(1):1–35, March 2014.

[Bue09] Francisco Buera. "A dynamic model of entrepreneurship with borrowing constraints: theory and evidence." *Annals of Finance*, **5**(3):443–464, June 2009.

[Cas80] J. Casti. "On the general inverse problem of optimal control theory." *Journal of Optimization Theory and Applications*, **32**(4):491–497, 1980.

[CD05] Marco Cagetti and Mariacristina De Nardi. "Entrepreneurship, frictions, and wealth." Working Paper Series WP-05-09, Federal Reserve Bank of Chicago, 2005.

[CDR03] Ana Castañeda, Javier Díaz-Giménez, and José-Víctor Ríos-Rull. "Accounting for the U.S. Earnings and Wealth Inequality." *Journal of Political Economy*, **111**(4):pp. 818–857, 2003.

[Cha88] Fwu-Ranq Chang. "The Inverse Optimal Problem: A Dynamic Programming Approach." *Econometrica*, **56**(1):pp. 147–172, 1988.

[Che07] Xiaohong Chen. "Large Sample Sieve Estimation of Semi-Nonparametric Models." In J.J. Heckman and E.E. Leamer, editors, *Handbook of Econometrics*, volume 6B, chapter 76. Elsevier, 1 edition, 2007.

[CJ93] Ricardo J. Caballero and Adam 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, May 1993.

[CJ04] A. Colin Cameron and Per Johansson. "Bivariate Count Data Regression Using Series Expansions: With Applications." Working Papers, University of California, Davis, Department of Economics 9815, University of California, Davis, Department of Economics, July 2004.

[CL14] Xiaohong Chen and Zhipeng Liao. "Sieve M inference on irregular parameters." *Journal of Econometrics*, **182**(1):70–86, 2014.

[CLS14] Xiaohong Chen, Zhipeng Liao, and Yixiao Sun. "Sieve inference on possibly misspecified semi-nonparametric time series models." *Journal of Econometrics*, **178**(P3):639–658, 2014.

[CT14] A. Colin Cameron and Pravin K. Trivedi. *Regression Analysis of Count Data.* 2 edition, 2014.

[Dav12] Joel David. "The Aggregate Implications of Mergers and Acquisitions." Technical report, 2012.

[EJ02] Jan Eeckhout and Boyan Jovanovic. "Knowledge Spillovers and Inequality." *American Economic Review, American Economic Association*, **92**(5):1290–1307, December 2002.

[ER06] Andrea L. Eisfeldt and Adriano A. Rampini. "Capital reallocation and liquidity." *Journal of Monetary Economics*, **53**(3):369–399, April 2006.

[Fel71] Willam Feller. *An Introduction to Probability Theory and Its Applications*, volume 2. John Wiley Sons, Inc., 1971.

[Gab09] Xavier Gabaix. "Power Laws in Economics and Finance." *Annuals Review of Economics*, 2009.

[GH86] Sanford J Grossman and Oliver D Hart. "The Costs and Benefits of Ownership: A Theory of Vertical and Lateral Integration." *Journal of Political Economy*, **94**(4):691–719, August 1986.

[GK03] Michelle Gittelman and Bruce Kogut. "Does good science lead to valuable knowledge? Biotechnology firms and the evolutionary logit of citation patterns." *Management Science*, pp. 366–382, 2003.

[GKO15] Fatih Guvenen, Fatih Karahan, Serdar Ozkan, and Jae Song. "What Do Data on Millions of U.S. Workers Reveal about Life-Cycle Earnings Risk?" NBER Working Papers 20913, National Bureau of Economic Research, Inc, January 2015.

[GLL] Olivier Guéant, Jean-Michel Lasry, and Pierre-Louis Lions. "Mean field games and applications.".

[GN87] A Ronald Gallant and Douglas W Nychka. "Semi-nonparametric Maximum Likelihood Estimation." *Econometrica, Econometric Society*, **55**(2):363–90, March 1987.

[Gri57] Zvi Griliches. "Hybrid Corn: An Exploration in the Economics of Technological Change." *Econometrica*, **25**(4):501–522, 1957.

[GRS98] Shiferaw Gurmu, Paul Rilstone, and Steven Stern. "Semiparametric estimation of count regression models1." *Journal of Econometrics, Elsevier*, **88**(1):123–150, November 1998.

[GS72] Iosif I. Gihman and Anatolij V. Skorohod. *Stochastic Differential Equations*. Springer-Verlag Berlin Heidelberg, 1972.

[GS14] Alberto Galasso and Mark Schankerman. "Patents and Cumulative Innovation: Causal Evidence from the Courts." NBER Working Papers 20269, National Bureau of Economic Research, Inc, July 2014.

[HJT98] Rebecca Henderson, Adam B. Jaffe, and Manuel Trajtenberg. "Universities As A Source Of Commercial Technology: A Detailed Analysis Of University Patenting, 1965-1988." *The Review of Economics and Statistics*, **80**(1):119–127, February 1998.

[HJT01] Bronwyn H Hall, Adam B Jaffe, and Manuel Trajtenberg. "The NBER Patent Citations Data File: Lessons, Insights and Methodological Tools." CEPR Discussion Papers 3094, C.E.P.R. Discussion Papers, December 2001.

[HP10] Gerard Hoberg and Gordon Phillips. "Product Market Synergies and Competition in Mergers and Acquisitions: A Text-Based Analysis." *Review of Financial Studies*, **23**(10):3773–3811, October 2010.

[Hug93] Mark Huggett. "The risk-free rate in heterogeneous-agent incomplete-insurance economies." *Journal of Economic Dynamics and Control*, **17**(5-6):953–969, 1993.

[Joh04] Michael Johannes. "The Statistical and Economic Role of Jumps in Continuous-Time Interest Rate Models." *The Journal of Finance*, **59**(1):pp. 227–260, 2004.

[JR86] Boyan Jovanovic and Rafael Rob. "The Diffusion and Growth of Knowledge." Working Papers, C.V. Starr Center for Applied Economics, New York University 86-20, C.V. Starr Center for Applied Economics, New York University, 1986.

[JR02] Boyan Jovanovic and Peter L. Rousseau. "The Q-Theory of Mergers." *American Economic Review*, **92**(2):198–204, May 2002.

[JTH93] Adam B Jaffe, Manuel Trajtenberg, and Rebecca Henderson. "Geographic Localization of Knowledge Spillovers as Evidenced by Patent Citations." *The Quarterly Journal of Economics*, **108**(3):577–98, August 1993.

[Kou02] S. G. Kou. "A Jump-Diffusion Model for Option Pricing." *Management Science*, **48**(8):pp. 1086–1101, 2002.

[KPS12] Leonid Kogan, Dimitris Papanikolaou, Amit Seru, and Noah Stoffman. "Technological Innovation,Resource Allocation and Growth." *working paper*, 2012.

[Lar82] Kenneth J. Singleton Lars Peter Hansen. "Generalized Instrumental Variables Estimation of Nonlinear Rational Expectations Models." *Econometrica*, **50**(5):1269–1286, 1982.

[LDY11] Ronald Lai, Alexander D'Amour, Amy Yu, Ye Sun, and Lee Fleming. "Disambiguation and Co-authorship Networks of the U.S. Patent Inventor Database (1975 - 2010)." *Research Policy*, 2011.

[Lin95] Bruce G. Lindsay. "Mixture Models: Theory, Geometry and Applications." *NSF-CBMS Regional Conference Series in Probability and Statistics*, **5**:pp. i–iii+v–ix+1–163, 1995.

[LM14] Robert Lucas and Benjamin Moll. "Knowledge Growth and the Allocation of Time." *Journal of Political Economy, University of Chicago Press*, **122**(1):1 − 51, 2014.

[LS99] Naomi R. Lamoreaux and Kenneth L. Sokoloff. "Inventors, Firms, and the Market for Technology in the Late Nineteenth and Early Twentieth Centuries." In *Learning by Doing in Markets, Firms, and Countries*, NBER Chapters, pp. 19–60. National Bureau of Economic Research, Inc, March 1999.

115

[Luc87] Robert E. Lucas. *Models of Business Cycles*. Wiley-Blackwell, 1987.

[Lut07] Erzo G.J. Luttmer. "On the Mechanics of Firm Growth." Working Papers, University of Minnesota, Department of Economics 2007-4, University of Minnesota, Department of Economics, 10 2007.

[Lut13] Erzo G.J. Luttmer. "Eventually, Noise and Imitation Implies Balanced Growth." 2013 Meeting Papers, Society for Economic Dynamics 91, 8 2013.

[Mat03] Rosa L. Matzkin. "Nonparametric Estimation of Nonadditive Random Functions." *Econometrica*, **71**(5):1339–1375, 2003.

[MKV15] Benjamin Moll, Greg Kaplan, and Gianluca Violante. "Monetary Policy According to HANK." 2015.

[Nev65] Jacques Neveu. *Mathematical Foundations of the Calculus of Probability*. HOLDEN-DAY, Inc., 1965.

[Pak86] Ariel S Pakes. "Patents as Options: Some Estimates of the Value of Holding European Patent Stocks." *Econometrica, Econometric Society*, **54**(4):755–84, July 1986.

[POB07] Ariel Pakes, Michael Ostrovsky, and Steven Berry. "Simple estimators for the parameters of discrete dynamic games (with entry/exit examples)." *RAND Journal of Economics*, **38**(2):373–399, 06 2007.

[PT14] Jesse Perla and Christopher Tonetti. "Equilibrium Imitation and Growth." *Journal of Political Economy*, **122**(1):52 – 76, 2014.

[Qua00] Vincenzo Quadrini. "Entrepreneurship, Saving and Social Mobility." *Review of Economic Dynamics*, **3**(1):1–40, January 2000.

[Rao92] Prakasa Rao. *Identifiability of Stochastic Models*. Academic Press, 1992.

[RC13] Michael Roach and Wesley M. Cohen. "Lens or Prism? Patent Citations as a Measure of Knowledge Flows from Public Research." *Management Science*, **59**(2):504–525, 2013.

[Ris78] H. Risken. *The Fokker-Planck Equation.* Springer-Verlag Berlin Heidelberg, 1978.

[RR08] Matthew Rhodes-Kropf and David T. Robinson. "The Market for Mergers and the Boundaries of the Firm." *Journal of Finance*, **63**(3):1169–1211, 06 2008.

[SA11] Jasjit Singh and Ajay Agrawal. "Recruiting for Ideas: How Firms Exploit the Prior Inventions of New Hires." *Management Science*, **57**(1):129–150, January 2011.

[Sch10] Paul Schrimpf. "Identification and estimation of dynamic games with continuous states and controls." 2010.

[Sch15] Lawrence D. W. Schmidt. "Climbing and Falling Off the Ladder: Asset Pricing Implications of Labor Market Event Risk." 2015.

[Ser10] Carlos J. Serrano. "The dynamics of the transfer and renewal of patents." *RAND Journal of Economics*, **41**(4):686–708, 2010.

[Ser14] Amit Seru. "Firm boundaries matter: Evidence from conglomerates and R&D activity." *Journal of Financial Economics*, **111**(2):381–405, 2014.

[Tei67] Henry Teicher. "Identifiability of Mixtures of Product Measures." *The Annals of Mathematical Statistics*, **38**(4), 1967.

[Wil10] Oliver E. Williamson. "Transaction Cost Economics: The Natural Progression." *American Economic Review*, **100**(3):673–90, June 2010.

[WTP13] Mike Waugh, Christopher Tonetti, and Jesse Perla. "Equilibrium Technology Diffusion, Trade, and Growth." 2013 Meeting Papers 484, Society for Economic Dynamics, 2013.

[XLG03] Wei Xu, Xin Liu, and Yihong Gong. "Document Clustering Based On Non-negative Matrix Factorization." 2003.

[ZDB98] Lynne G Zucker, Michael R Darby, and Marilynn B Brewer. "Intellectual Human Capital and the Birth of U.S. Biotechnology Enterprises." *American Economic Review*, **88**(1):290–306, March 1998.

[ZDF11] Lynne G. Zucker, Michael R. Darby, and Jason Fong. "Communitywide Database Designs for Tracking Innovation Impact: COMETS, STARS and Nanobank." Working Paper 17404, National Bureau of Economic Research, September 2011.

[Zha14] Xiaolan Zhang. "Who Bears Firm-Level Risk? Implications for Cash Flow Volatility." 2014.